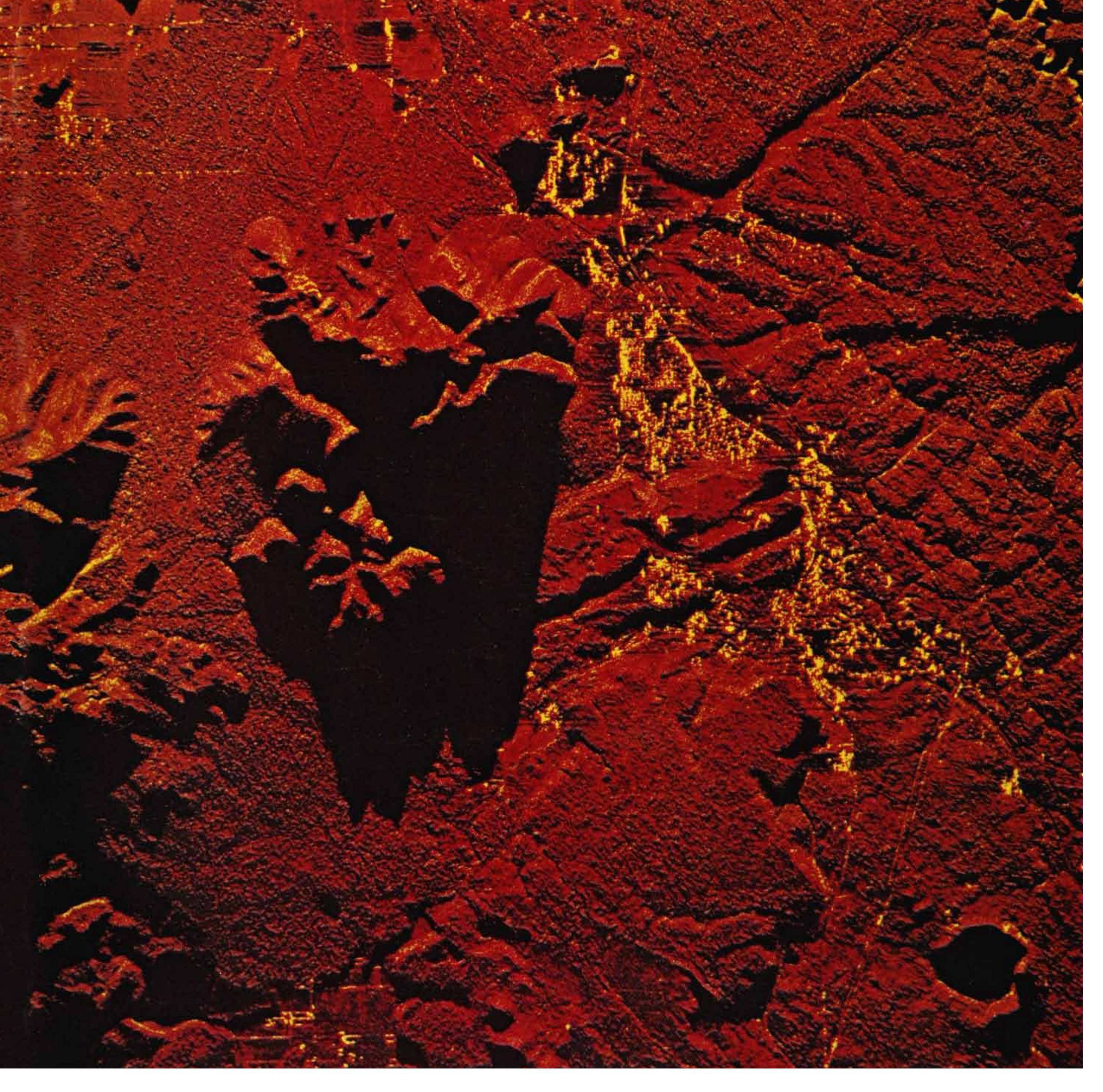
INVESTIGACION Y CIENCIA

Edición en español de **SCIENTIFIC AMERICAN**



RADAR DE IMAGENES LATERALES

Diciembre 1977

140 PTAS.

Los espacios en gris
corresponden a publicidad
en la edición impresa

Número 15
Diciembre 1977

INVESTIGACION Y
CIENCIA
Edición en español de SCIENTIFIC AMERICAN

EDITA Prensa Científica, S.A.
Calabria, 235-239
Barcelona-15
ESPAÑA

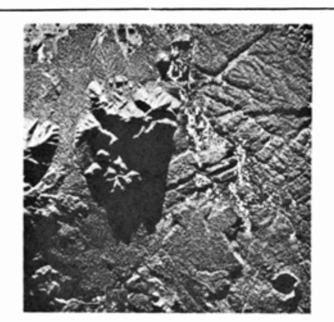
Colaboradores de este número:

Asesoramiento y traducción:

Ramón Canal: *Estrellas de rayos X en cúmulos globulares;* Ramón Pascual: *Partículas fundamentales con encanto;* Amando García: *Radar de imágenes laterales aerotransportado;* José Luis Rodríguez: *Estructura y funcionamiento de los antígenos histocompatibles;* Luis Bou: *La solución del problema del mapa de cuatro colores;* Nuria Rafel: *Cómo empezó la Edad del Hierro;* Silvia López: *Alucinaciones;* Luis Bou: *Juegos matemáticos;* José E. Myro y Alicia Menéndez: *Taller y laboratorio;* Gerardo Espinosa: *Ciencia y sociedad.*

Libros:

Luis Alonso



LA PORTADA

La ilustración de la portada parece una fotografía aérea, pero no lo es. Es una imagen de la zona de Flagstaff, Arizona, realizada por un sistema de radar lateral aerotransportado. La antena de dicho sistema se fija en el fuselaje del avión y apunta lateralmente. Los impulsos de microondas emitidos por la antena inciden en el suelo en ángulo agudo. Las reflexiones de los impulsos los recibe la antena, aparecen en la pantalla de un tubo de rayos catódicos y se registran sobre una película fotográfica, formando un holograma radárico. Luego se crea la imagen iluminando el holograma con luz monocromática y se registra dicha imagen reconstruida en película de color (véase "Radar de imágenes laterales aerotransportado", de Homer Jensen, L.C. Graham, Leonard J. Porcello y Emmett N. Leith, en este número). El avión, en este caso, volaba a 40 kilómetros al este de Flagstaff, yendo de izquierda a derecha (de norte a sur) a una altitud de 12.200 metros. La prominencia situada a la izquierda es el monte Elden. La constelación de manchitas amarillas que se extienden desde arriba al centro, hacia el ángulo inferior derecho es Flagstaff, y la especie de cráter visible en dicho ángulo es el Lago Seco. La línea de trazos amarillos de la parte superior izquierda es una serie de reflexiones radáricas de una línea de alta tensión.

Suscripciones:

Prensa Científica, S.A.
Departamento de suscripciones
Calabria, 235-239
Barcelona-15 (España)

Distribución para España:

Distribuciones de Enlace, S.A.
Ausias March, 49
Barcelona-10 (España)

Distribución para los restantes países:

Editorial Labor, S.A.
Calabria, 235-239
Barcelona-15 (España)

Precio de venta de este número:

España: 140 pesetas
Extranjero: 2 U.S. $

Condiciones de suscripción:

España:
Un año (12 números): 1.540 pesetas
Extranjero:
Un año (12 números): 30 U.S. $

Ejemplar atrasado ordinario:
160 pesetas

Ejemplar atrasado extraordinario:
225 pesetas.

Solicitado control **OJD**

PROCEDENCIA DE LAS ILUSTRACIONES

La fotografía de la portada es cortesía de Goodyear Aerospace Corporation

Página	Fuente
7	Sodi-Pallares
8-11	Sodi-Pallares y A. Sellés
12	Sodi-Pallares y A. Sellés *(arriba);* Sodi-Pallares *(abajo)*
13-14	Sodi-Pallares y A. Sellés
17	© National Geographic Society-Palomar Observatory Sky Survey
18	Cerro Tololo Inter-American Observatory
19-27	Lorelle M. Raboni
33	Jon Brenneis
34-48	Allen Beechel
50	Paul B. MacCready
53	Goodyear Aerospace Corporation and Aero Service Division, Western Geophysical Company of America
54-56	George V. Kelvin
57	Goodyear Aerospace Corporation and Aero Service Division, Western Geophysical Company of America
58-61	George V. Kelvin
62-63	Goodyear Aerospace Corporation and Aero Service Division, Western Geophysical Company of America *(arriba, izquierda);* George V. Kelvin *(arriba, derecha);* Environmental Research Institute of Michigan *(abajo);*
72-73	Ben Rose
74-76	Ilil Arbel
79-90	Lorelle M. Raboni
93	Y. K. Rao, Universidad de Washington
94	Tom Prentiss *(arriba),* Type Systems Inc. *(abajo)*
95-97	Robert Maddin, Universidad de Pennsylvania
98	Type Systems Inc.
101-103	Ronald K. Siegel, Universidad de California en Los Angeles
104-106	Alan D. Iselin
116	Michael Goodman
117	Michael Goodman *(arriba y centro);* Ben Rose *(abajo)*
118-120	Michael Goodman

Dep. legal: B. 38.999-76
Fotocomposición: Fort, S.A.
Rosellón, 33 - Barcelona-15
Cayfosa. Santa Perpetua de Moguda. Barcelona.

Printed in Spain - Impreso en España

Cartas

Señor Director:

El artículo "La eliminación de los residuos radiactivos" de Bernard L. Cohen [INVESTIGACIÓN Y CIENCIA, agosto], contiene errores graves y afirmaciones que pueden inducir a error. Por razones de espacio limitaré mi análisis a unas pocas faltas que parecen importantes.

El comentario de Cohen (páginas 8-9) respecto del difundido uso actual del plutonio como combustible en los reactores de agua ligera es cierto, pero probablemente es engañoso para el lector incauto. La controversia concerniente al reciclaje del plutonio tiene escasa relación con el plutonio instalado seguramente dentro del núcleo de un reactor en funcionamiento. Se ocupa únicamente del destino del plutonio una vez extraído del reactor.

Las observaciones intercaladas sobre la "hipótesis de linealidad" (páginas 9 y 10) suponen que si se abandona dicha hipótesis, se la reemplazará con un modelo de umbral o algún otro de menor toxicidad. Pero si llegan a tener una aceptación general unas hipótesis tales como la de Karl Z. Morgan ("Disminución sugerida de la exposición admisible al plutonio y a otros elementos transuránicos", *American Industrial Hygiene Association Journal*, páginas 567-575; agosto, 1975) referente a isótopos con actividad alfa, el abandono de la hipótesis de linealidad podría conducir a un aumento de la magnitud estimada de los riesgos sanitarios de origen nuclear, y no a "rebajarlos drásticamente".

Cohen debe fundamentar su aserción (página 15) de que "cerámicas u otros materiales más resistentes" podrán mantener su estabilidad y escasa lixiviación durante largo tiempo mientras estén sometidos a calentamiento térmico, intensa radiación, grandes presiones de la montera y a la salmuera circundante.

La exposición de Cohen sobre procesos de intercambio iónico que tienden a retardar el traslado de desechos radiactivos a través de acuíferos (página 15) simplifica en exceso el asunto. Ya fue criticada por F. von Hippel, D. G. Jacobs y J. E. Turner en *Physics Today* (páginas 68-69, agosto de 1976, y páginas 15 y 86, noviembre de 1976).

La probabilidad calculada de la liberación de desechos (página 15, final) se basa en suponer una "misma probabilidad de escape para los átomos de radio y de los residuos". Esta suposición es infundada, pues es posible imaginar situaciones de falla de depósitos de desechos que no son aplicables al radio existente en la tierra. Por ejemplo, (1) existe la posibilidad de que la dilatación térmica provocada por los desechos o el hundimiento de la mina tras el relleno pudiese fracturar las formaciones de pizarra suprayacentes (ilustradas en la página 10), trayendo por consecuencia una imprevista intrusión de agua subterránea (véase W. Hambleton "El problema no resuelto de los desechos nucleares", *Technology Review*, páginas 15-19; marzo-abril 1972) y (2) el radio existente en la tierra no tiene uno o más socavones de mina "sellados" que conducen hasta él, y se ha trabajado poco sobre la factibilidad de crear obturadores estancos de larga duración. Además, el saber que New México es árido en la actualidad es útil por lo que respecta a los breves períodos de semidesintegración del estroncio 90 y del cesio 137, pero no viene mucho al caso de tales períodos del iodo 129 o de los actínidos (más de 24.000 años). El no abordar tales hechos quita todo sentido al cálculo de probabilidades realizado por Cohen.

La creencia del autor (página 17) de que la fisión nuclear elimina la radiactividad de la tierra es incompleta. Omite mencionar que depósitos de uranio que yacían (más o menos) seguros, bajo tierra, están siendo excavados a fin de fisionarlos. Al hacerlo, se deja expuesto a la intemperie radio, radón y sus descendientes en los montículos de relaves a nivel de la superficie. Partiendo de la hipótesis de linealidad y de una supuesta población global de 4000 millones, se puede calcular que cada año de funcionamiento de un reactor provoca con el tiempo más de 300 muertes por cáncer latente por esta sola causa.

En la comparación entre riesgos químicos y radiactivos (página 17) debería hacerse notar que las toxinas químicas pueden neutralizarse, si se desea, mediante transformaciones químicas adecuadas. Las toxinas radiológicas siguen siendo peligrosas sea cual fuere su forma química, siempre que pueda postularse algún mecanismo verosímil para "administrar" una dosis. Además, la toxicidad absoluta indicada para desechos nucleares después de 100 años (10^7 dosis letales) contradice la información presentada en los gráficos de las páginas 14 y 15 (10^{10} dosis letales).

El examen de la probabilidad y naturaleza de diversas actividades humanas futuras (páginas 19) es enteramente especulativo. Sin embargo, me permito observar que "unas perforaciones de prospección realizadas al azar" quizá no sean una base útil para la predicción. Suponiendo cierta continuidad en la naturaleza humana, la curiosidad seguirá acompañándonos por largo tiempo. De modo que las pruebas de que en el pasado se ha perforado (que serían obvias si se dejan señales como advertencia) podrían muy bien aumentar la probabilidad de actividad humana en dicho sitio en el futuro...

Respecto de la declaración final del autor, es improbable que en el futuro haya ninguna fuente "barata" de energía de alta calidad.

Cohen omite además la valoración de los riesgos relacionados con las operaciones de transporte y manipulación anteriores al entierro de los desechos.

Es claro que tenemos desechos radiactivos y deberíamos preocuparnos de colocarlos en algún sitio en donde sea improbable que nos hagan daño (ni a las generaciones futuras). Personalmente, preferiría almacenar los desechos de alto nivel o el combustible agotado y no reprocesado en una forma fácilmente recuperable en una estructura construida y situada en lo profundo de una mina de sal u otra formación geológica relativamente estable, mientras se procede a una evaluación más cabal de otros sistemas de eliminación de los que (a menos que surja una solución conceptual imprevista) puedan conseguirse en los próximos veinte años.

RICHARD SCLOVE

Departamento de Ingeniería Nuclear
Instituto de Tecnología de Massachusetts
Cambridge

Señor Director:

En mi respuesta a la carta del Sr. Sclove seguiré el orden de sus comentarios.

Mi artículo se basó en un ensayo más detallado, aparecido en *Reviews of Modern Physics* (enero, 1977), una revista reputada y minuciosamente criticada (mi artículo fue revisado por 14 personas). Si el artículo contiene "graves errores", el Sr. Sclove debería presentar un estudio ante dicha revista.

Pensé que mi exposición, hecha en

una frase, acerca del quemado del plutonio, era interesante y no veo cómo pueda considerársela engañosa. Hay quienes parecen creer que cualquier uso del plutonio, y hasta su existencia misma, son pecaminosos.

La posición de que es mucho más probable que la "hipótesis de linealidad" sobrestime y que no subestime los efectos nocivos de dosis pequeñas de irradiación está avalada por todos los grupos oficiales de seguridad radiológica, incluso por el comité sobre los efectos biológicos de la radiación ionizante de la National Academy of Sciences (Informe BEIR), la Comisión Nacional estadounidense sobre Protección Radiológica y Medición de la Radiación (NCRP, Publicación N.º 43), y el comité científico sobre los efectos de la radiación atómica, de las Naciones Unidas *(Ionizing Radiation: Levels and Effects)*. La proposición de que pudiera no ser así, a la que se refiere el señor Sclove, fue rechazada por la Comisión Internacional sobre Protección Radiológica en su decisión de no rebajar las exposiciones admisibles. En vista de tales circunstancias creo que mi observación, intercalada en una frase, estaba justificada. En todo caso, no formaba parte de mi análisis.

En mi artículo no hay una afirmación acerca de la cerámica como la que menciona el Sr. Sclove en su carta. Mi comentario sobre la cerámica se basa en la labor de G. McCarthy y Rustum Roy, de la Pennsylvania State University.

La crítica de von Hippel, aparecida en *Physics Today*, sobre mi tratamiento de los efectos del intercambio iónico en el transporte de desechos radiactivos, fue respondida a cabalidad en mi respuesta a su carta (publicada en el número siguiente de dicha revista). Por entonces von Hippel escribió que continuaba estudiando el problema, y aunque me ha escrito al menos un par de veces desde entonces sobre otros temas, no ha vuelto a mencionar ese problema. La carta de Jacobs-Turner era básicamente un tratamiento más detallado que el mío más que una crítica del mío, y hay pocos desacuerdos entre nosotros, según se explicaba en mi respuesta, publicada conjuntamente con la carta de ellos.

La fractura de la pizarra suprayacente por problemas térmicos o de hundimiento fue estudiada por el grupo de Oak Ridge y no se la considera un problema grave. Se está trabajando en el problema de la obturación de los socavones y no se cree que sea notablemente difícil. El área del socavón es minúscula, en la mayoría de los proyectos está bastante lejos de los desechos enterrados y la sal se obtura por sí misma. Además *hav* agujeros en el terreno (por ejemplo, cuevas, minas y fisuras) que contienen el radio con el que se comparan los desechos. Esto, para no hablar de la comparación entre el radio y los desechos con respecto al agua subterránea, los ríos, vientos, animales, ciclos de hielo-deshielo, vegetación, que actúan en las superficies, etcétera. En mi estudio no se supone que el lugar del entierro sea árido o vaya a seguir siéndolo. No concuerdo en que haya alguna injusticia en que yo compare las probabilidades de escape de los desechos enterrados y del radio existente en el terreno por encima de los desechos.

El asunto de los relaves de la minería y molienda del uranio es más bien un tema aparte que se estudia en otros ensayos míos (por ejemplo, *Bulletin of the Atomic Scientists*, febrero 1976, página 61). Se está trabajando activamente en ese problema y resolverlo no parece ni muy difícil ni muy costoso. De paso, con el mismo cálculo que según cita el señor Sclove indica 300 muertes eventuales por cáncer por dicha causa, se deduce que por el hecho de quemar uranio se evitan 5000 muertes por cáncer.

La toxicidad química del arsénico y del bario *no* se ve prácticamente afectada por cambios corrientes de su forma química. En la ilustración de la página 16 había un error en cuanto que "(después de 100 años)" debía leerse "(después de 500 años)". Los números correctos aparecen en el texto y en las ilustraciones de las páginas 14 y 15.

No entiendo la objeción del señor Sclove a mi tratamiento del problema de la perforación, ni simpatizo con ella. ¿Puede sugerir él un tratamiento más apropiado? De las perforaciones que se realicen en un futuro lejano no puede provenir nada espantoso.

No veo razón por qué la energía nuclear no será más barata en el futuro, y mantenerse así durante milenios.

En otros ensayos se ha estudiado los peligros que para el público presentan el transporte y la manipulación de desechos y se ha descubierto que tienen una importancia mucho menor que los aquí analizados. Se estima que las liberaciones de radiactividad a partir de accidentes de transporte ocasionarían un promedio de mucho menos de 0,01 muertes por cada año de energía eléctrica enteramente nuclear.

Es difícil reñir con el señor Sclove en su petición de más investigaciones, y la labor investigadora se está acelerando. No obstante, también me cuesta entender la intensidad de su preocupación por los desechos de la energía nuclear, que jamás han dañado a nadie y que, según se demuestra en mi artículo, probablemente no ocasionarán nunca siquiera una eventual muerte por año, cuando estamos matando por lo menos a 10.000 estadounidenses cada año con los desechos, contaminadores atmosféricos, de la combustión de la hulla, única fuente alternativa viable de energía eléctrica. Además, no puedo comprender su insistencia en que se parta de los supuestos más prudentes y se tenga en cuenta los sucesos más improbables al evaluar los peligros de los residuos nucleares, cuando se adopta la filosofía contraria al evaluar los peligros de los residuos de la combustión de la hulla y todos los demás riesgos existentes en nuestra sociedad. En la literatura científica hay estudios en los que se estima que las muertes anuales por la contaminación atmosférica ocasionada por la combustión de la hulla son varias veces 10.000. Dichos estudios jamás han sido refutados. Asimismo los efectos de la radiación son mucho mejor comprendidos que los de la contaminación atmosférica, y los niveles de radiación examinados en este caso son una fracción minúscula de la radiación natural de fondo, mientras que en el caso de la contaminación atmosférica los niveles son varias veces superiores al trasfondo natural ¿Es demasiado pedir que haya cierto equilibrio y perspectiva al considerar nuestros problemas de eliminación de residuos?

BERNARD L. COHEN

Departamento de Física y Astronomía
Universidad de Pittsburgh
Pittsburgh, Pennsylvania

ERRATA

Número 11, agosto de 1977

Pág.	*columna*	*línea*	*dice*	*debe decir*
3	3	32 y ss.		WILLIAM D. LILLEY Consulting geologist Elsmere, N. Y.

Los autores

DEMETRIO SODI-PALLARES y MIGUEL GONZALEZ AHUMADA ("El tratamiento metabólico de las enfermedades del corazón") trabajan en el Instituto Nacional de Cardiología, de México, que preside desde 1944 el primero. Sodi-Pallares nació en la capital azteca en 1913, en cuya Universidad Nacional se graduó en Medicina y Cirugía en 1936. Siguió cursos de especialidad en cardiología en la Michigan University. En 1944 se hizo cargo del Departamento del Hospital General. Desde 1946 enseña en la Universidad Nacional Autónoma. Miembro honorario de múltiples sociedades médicas americanas y europeas, ha publicado numerosas obras sobre cardiología. González Ahumada se formó en la Escuela Médico Militar, ha pasado años enseñando en la Universidad Nacional Autónoma y actualmente dirige la secretaría del Capítulo Mexicano de la Société Internationale de Medicine Humaniste Néo-Hippocratique.

GEORGE W. CLARK ("Estrellas de rayos X en cúmulos globulares") es profesor de física en el Instituto de Tecnología de Massachusetts. Estudió en el Harvard College y en el M.I.T., en donde obtuvo en 1952 su doctorado en física y en el que ha permanecido desde entonces. La relación de Clark con la astronomía de rayos X arranca de 1959, época en la que trabajó con Riccardo Giacconi en la detección de fuentes celestes de rayos X distintas del sol. Esos estudios fueron seguidos por experimentos en cohetes, que culminaron con el descubrimiento, en 1962, de la primera estrella de rayos X, Sco X-1, por Giacconi y su equipo. Clark empezó sus propias investigaciones en astronomía de rayos X en 1964, utilizando detectores instalados primero en globos y más tarde en satélites. Dirige el laboratorio de rayos X del Third Small Astronomical Satellite (SAS-3), que opera día y noche controlado desde el M.I.T.

ROY F. SCHWITTERS ("Partículas fundamentales con encanto") es profesor asociado de física en la Stanford University; trabaja en el Stanford Linear Accelerator Center (SLAC). Natural de Seattle, Washington, estudió en el Instituto de Tecnología de Massachusetts, en donde estuvo durante nueve años "muy agradables" realizando trabajos en física. Tras doctorarse en 1971, se trasladó a California como colaborador de investigación del SLAC, en cuyo puesto ayudó a diseñar y construir el detector magnético del proyecto SPEAR.

HOMER JENSEN, L. C. GRAHAM, LEONARD J. PORCELLO y EMMETT N. LEITH ("Radar de imágenes laterales aerotransportado") están interesados, desde hace tiempo, en los sistemas de radar. Jensen es el vicepresidente encargado de los reconocimientos con radar de la compañía Aero Services, una filial de la Litton Industries. Ha estado relacionado con los sistemas de percepción remota aerotransportados, habiendo participado en las primeras experiencias con magnetómetros aerotransportados, que se vieron coronadas por el éxito. Graham es director de investigación y desarrollo de la División de Arizona de la Goodyear Aerospace Corporation. Estudió en la Universidad de Colorado, en Boulder, donde recibió su título de licenciado en ingeniería en 1950. A continuación, se incorporó a los Laboratorios Bell, y obtuvo el diploma de graduado en el Instituto de Tecnología Stevens; ingresó en la Goodyear en 1956. Porcello es actualmente director de la Tucson Arizona, una filial de la empresa Science Applications, Inc. En 1963 recibió el título de doctor en ingeniería eléctrica en la Universidad de Michigan. Desde 1967 a 1972 fue director del laboratorio de radar y electrónica de esta Universidad. En 1976 inició su presente trabajo. Finalmente, Leith es director científico de los Willow Run Laboratories de Michigan. Es más conocido por sus actividades en el campo de la holografía.

BRUCE A. CUNNINGHAM ("La estructura y función de los antígenos histocompatibles") es profesor adjunto de bioquímica en la Rockefeller University, de Nueva York. Se graduó en química por la Universidad de Dubuque, doctorándose en bioquímica por la Yale University en 1966. Trabajó durante dos años como postgraduado en el laboratorio de Gerald M. Edelman, en la Rockefeller; entró en el claustro de la facultad como profesor asistente, en 1968. El interés principal de Cunningham ha sido siempre la química de proteínas, que ha aplicado a la determinación de la estructura de las moléculas de anticuerpos.

KENNETH APPEL y WOLFGANG HAKEN ("La solución del problema del mapa de los cuatro colores") son profesores de matemáticas de la Universidad de Illinois. Appel se formó en el Queens College y en la Universidad de Michigan, en donde se doctoró en 1959. Tras dos años en el equipo técnico del Institute for Defense Analysis, en Princeton, pasó al claustro de Illinois en 1961. A pesar de que la mayor parte de su trabajo guarda relación con problemas combinatorios en lógica y en teoría de grupos, ha utilizado los ordenadores para resolver problemas combinatorios desde hace más de veinte años. Haken, nacido en Berlín, recibió su grado de doctor en la Universidad de Kiel en 1953. En 1961 publicó una solución del "problema de los nudos", y al año siguiente le fue ofrecida una estancia en calidad de profesor visitante de matemáticas en Illinois. Desde 1963 hasta 1965 fue miembro interino del Instituto for Advanced Study, en Princeton, New Jersey.

ROBERT MADDIN, JAMES D. MUHLY y TAMARA S. WHEELER ("Cómo empezó la Edad del Hierro") forman un grupo interdisciplinario de arqueólogos y metalúrgicos de la Universidad de Pennsylvania. Maddin es profesor de metalurgia y miembro del Applied Science Center for Archaeology. Se licenció en ingeniería en la Universidad de Yale, recibiendo el doctorado en 1948. En 1955 pasó a la Universidad de Pennsylvania. Muhly es profesor adjunto de historia antigua y editor de *Expedition,* revista que edita el museo de la Universidad. Se educó en las Universidades de Minnesota y Yale, doctorándose por la última en 1969. Wheeler es profesora ayudante de metalurgia y ciencia de los materiales. Asistió al Colegio Bryn Mawr, en donde alcanzó el doctorado en arqueología en 1973. Después de impartir un curso en el Swarthmore College de Atenas, pasó a Pennsylvania en 1974.

RONALD K. SIEGEL ("Alucinaciones") es psicólogo investigador en el departamento de Psiquiatría y Ciencias de la Conducta de la Universidad de California (Los Angeles), en donde enseña también. Realizó sus estudios en la Universidad de Brandeis, y obtuvo el doctorado en la Universidad de Dalhousie, en 1970. Antes de entrar en la UCLA, asistió durante dos años como becario posgraduado en el departamento de farmacología Albert Einstein College of Medicine, de Nueva York.

El tratamiento metabólico de las enfermedades del corazón

El corazón es un sistema biológico abierto, que intercambia materia y energía con el medio que lo rodea. La terapéutica a seguir para el tratamiento de sus trastornos debe ser asimismo termodinámica, es decir, de equilibrio entre materia y energía

Demetrio Sodi-Pallares y Miguel González Ahumada

Albert Einstein, en su libro *La física, aventura del pensamiento* ("The Evolution of Physics"), describe las tres formas de energía conocidas, utilizando como ejemplo el juego mecánico de la montaña rusa. Distingue la energía potencial, la energía cinética y la energía que se pierde en forma de calor por rozamiento.

En la montaña rusa, cuando el vehículo de pasajeros se encuentra en la parte más alta del sistema, antes de iniciar su recorrido (es decir, con velocidad cero), dicho vehículo tiene su máxima energía potencial, pero carece de energía cinética. Cuando empieza a rodar cuesta abajo, su velocidad aumenta, y el vehículo adquiere máxima energía cinética en la parte más baja del sistema, por ser allí la velocidad mayor. En cambio, no hay, para entonces, energía potencial.

En todos los puntos intermedios, entre la cima o parte más alta y el valle o parte más baja del sistema, en donde hay velocidad variable y cierta elevación, el vehículo tiene energía potencial y energía cinética. La energía potencial aumenta y la cinética disminuye en el tramo de ascenso hasta la cima, y viceversa, en el tramo descendiente desde la cima hasta el valle, la energía cinética aumenta y la potencial disminuye.

Si estas dos formas de energía fueran las únicas, su suma daría siempre el mismo resultado, suma que podría llamarse "constante de movimiento".

En una montaña rusa ideal, el vehículo, después de bajar, podría subir a una altura similar al punto de salida. Pero, en la realidad, el vehículo de pasajeros nunca alcanza un punto tan alto como el de salida. Además, la suma de las energías potencial y cinética no permanece constante, sino que cada vez es menor; ello obedece a la aparición de una nueva forma de energía, constituida por el calor producido por el rozamiento de las ruedas del móvil sobre los rieles. La cantidad de calor producida corresponde a la disminución en energía mecánica; es decir, en las energías potencial y cinética. Quizás, como sugiere Einstein, la suma de las tres formas de energía a saber, calor, energía potencial y energía cinética, sí permanece constante. Según las propias palabras de Einstein: "En efecto, además de las energías cinética y potencial, nos encontramos también con el calor creado por rozamiento. ¿Corresponderá también este calor a la disminución de la energía mecánica, es decir, cinética y potencial? Si el calor puede considerarse como una forma de la energía, podría ser que la suma del calor, la energía cinética y la potencial, permanezca constante. No solamente el calor, sino que éste y otras formas de la energía tomadas en conjunto se comportan como una substancia, resultando indestructible su suma".

Las tres formas einstenianas de energía en el corazón

El estudio de las relaciones entre el calor y la energía mecánica compete a la termodinámica; su ámbito de investigación son los agregados de materia y energía que se denominan "sistemas". El corazón, por ejemplo, es un sistema biológico. Los sistemas más importantes que examina la termodinámica clásica son: el sistema aislado, el sistema cerrado y el sistema abierto. El sistema aislado no tiene ninguna relación con el medio que lo rodea y no intercambia materia ni energía con él; el sistema cerrado no intercambia materia, pero sí puede intercambiar libremente energía con su ambiente; por último, el sistema abierto puede intercambiar materia y energía.

El corazón es un sistema biológico que intercambia materia y energía con el medio que lo rodea. Desde el punto de vista termodinámico es, por tanto, un sistema abierto, en permanente comunicación con todos los órganos y tejidos del cuerpo humano, así como con el medio que lo rodea.

En el corazón se pueden considerar las tres formas de energía descritas por Einstein.

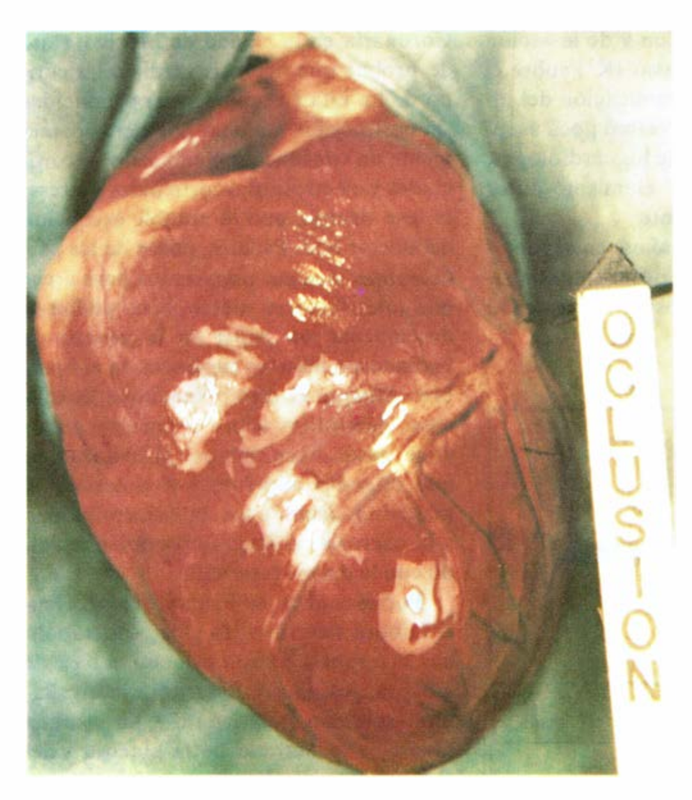
La energía cinética, conocida como "energía libre" en los sistemas biológicos. El corazón requiere de esta energía libre para su contracción y para las principales funciones del órgano. El adenosintrifosfato (ATP) es la molécula más importante por su alto contenido en energía libre; en efecto, una molécula de adenosintrifosfato ofrece alrededor de 7600 a 7800 calorías de energía libre. La fosfocreatina es otra importante molécula por su alto contenido en energía libre; sin embargo, no se presenta directamente al corazón para sus requerimientos, sino que se transforma en ATP de acuerdo con la reacción siguiente:

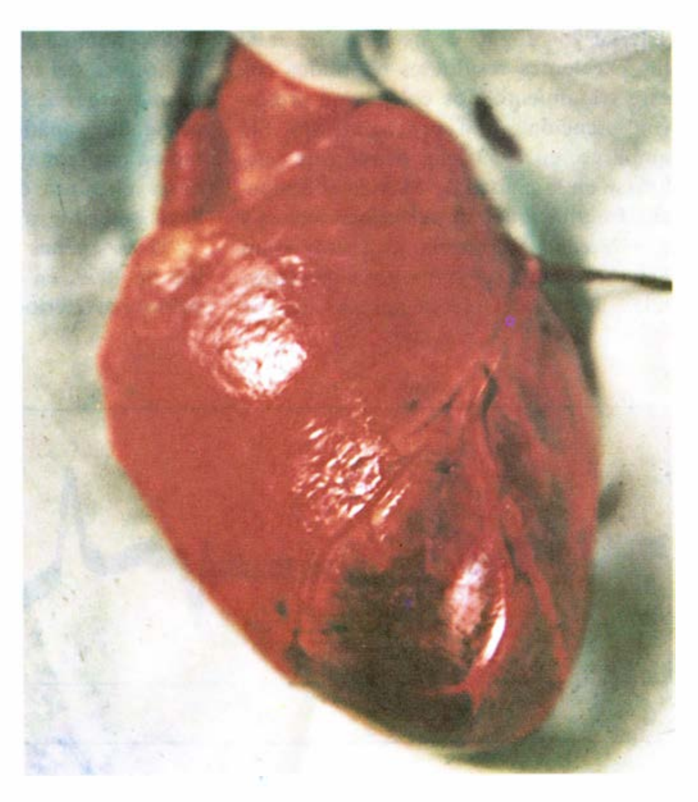
$$\text{Fosfocreatina} + \text{ADP} \rightleftarrows \text{Creatina} + \text{ATP}.$$

Esta reacción es reversible, pero en condiciones de equilibrio favorece la formación de ATP.

En un sentido metafórico, podríamos decir que la molécula de ATP es indispensable para mantener la vida de un sistema termodinámico abierto como es el corazón.

La energía potencial, que está representada en el corazón, sobre todo, por azúcar (el glucógeno) y una clase de grasas (los triglicéridos), que almacena el citoplasma de las fibras miocárdicas. Cuando el corazón trabaja activamente,





DIFERENCIA EN EL COLOR de la superficie epicárdica de un corazón bajo el efecto de la solución polarizante *(izquierda)* con otro corazón dejado a la evolución propia del infarto *(derecha)*. El corazón es un músculo hueco que por contracción de su pared muscular expulsa la sangre de su interior; la pared cardíaca consta de tres capas: endocardio o membrana interna, miocardio o musculatura cardíaca propiamente dicha y epicardio, que es la membrana externa o túnica serosa. Estamos en una situación de infarto cuando se forma un coágulo en una rama de las arterias coronarias que suministra sangre al corazón originando la obstrucción o cierre de la arteria y mortificando la zona afectada.

o está sometido a condiciones de emergencia (ejercicio, fiebre, estados angustiosos, etcétera), la energía potencial almacenada en el glucógeno y los triglicéridos se convierte en energía libre (energía cinética de Einstein) a través de las vías metabólicas (glicólisis anaeróbica, glicólisis aeróbica o ciclo de Krebs y fosforilación oxidativa). Se puede asemejar este cambio de energía potencial en energía libre al descrito por Einstein en la montaña rusa, de energía potencial a energía cinética.

La energía que no puede realizar trabajo. De hecho, esta afirmación es una expresión de la segunda ley de la termodinámica. Esta, en el enunciado de Kelvin, dice que: "es imposible construir un dispositivo que pueda, sin ningún otro efecto, levantar un objeto extrayendo energía térmica de otro". Hace énfasis en la limitación que existe para convertir calor en trabajo (entropía). Por extensión, se considera también a la entropía como la desorganización del sistema. Por ejemplo, en el caso de la montaña rusa, si el rozamiento que se produce entre las ruedas del vehículo y los rieles aumenta considerablemente, el sistema montaña rusa no trabaja, o lo que es lo mismo, se desorganiza.

La terapéutica metabólica, podría llamarse también termodinámica, que proponemos para numerosas enfermedades del corazón, tiene los siguientes propósitos: 1) Mantener en cantidades satisfactorias la energía potencial. Esta, por ejemplo, disminuye de modo importante en la isquemia miocárdica, que es la detención de la circulación arterial. Se ha visto, además, que la disminución del glucógeno muscular corre pareja con el daño miocárdico producido por la isquemia. 2) Mantener en cantidades adecuadas de energía libre, o, lo que es lo mismo, el número de moléculas sintetizadas del nucleótido adenosintrifosfato, procedente de las vías mencionadas. En la isquemia miocárdica también disminuye la energía libre y la cantidad de ATP baja notablemente, sobre todo a nivel del músculo próximo al endocardio del ventrículo izquierdo. 3) Evitar que la entropía aumente. Sabemos que en el sistema biológico corazón existe entropía, como en todo sistema; empero en las condiciones patológicas como es la isquemia miocárdica a que nos hemos referido, la entropía aumenta. El apartado 3) es una consecuencia lógica de los dos anteriores, ya que todo aumento de entropía significa una disminución de la energía potencial o de la energía libre, puesto que como ya dijimos, y de acuerdo con Einstein: H (energía potencial) G (energía libre) + S (entropía) $= K$ (constante). Si la entropía aumenta ($\Delta S\uparrow$), obligadamente el incremento de H (ΔH) o el incremento de G (ΔG) es negativo.

Historia del tratamiento metabólico

En el año de 1944, cierto familiar de uno de los autores (Sodi), de 64 años de edad, sufría de grave insuficiencia cardíaca global. El tratamiento prescrito por los cardiólogos era a base de digitálicos, diuréticos mercuriales y una dieta baja en colesterol. Cada vez que la paciente recibía el diurético, mostraba gran debilidad y quedaba exhausta. Sodi, que por esos años estudiaba medicina, preguntó la razón por la que se prescribían los diuréticos. La respuesta de los cardiólogos fue que se daban para eliminar la sal; entonces, preguntó, ¿por qué se la permite comer alimentos salados? La respuesta fue contundente: porque la dieta baja en colesterol, que se opone a la dieta sin sal, es necesaria para evitar el avance de la aterosclerosis generalizada. Insatisfecho con la respuesta, Sodi confeccionó una dieta baja en sodio y alta en potasio que permitió a la paciente dejar los diuréticos y una sobrevida, prácticamente normal, de 15 años.

Especializado ya en cardiología, Sodi empezó a prescribir la misma dieta en otras condiciones, como la hipertensión

arterial y la angina de pecho. A pesar de los excelentes resultados obtenidos con este tratamiento, los colegas no prestaron atención a tal procedimiento. Sin embargo, en 1959, con ocasión de celebrarse en la ciudad de México el Simposio Internacional de Arterioesclerosis y Enfermedad Coronaria, el doctor Hans Selye presentó sus hallazgos experimentales acerca del efecto agresor del ion sodio (Na^+) en el corazón y de la acción protectora del ion potasio (K^+) sobre el mismo órgano. La comunicación del investigador canadiense recibió poca aceptación por la mayoría de los cardiólogos, hasta el punto de que el mismo autor llegó a escribir lo siguiente:

"Hace apenas ocho años, en septiembre de 1959, en el Simposio Internacional sobre Arterioesclerosis y Enfermedad Coronaria en la ciudad de México, traté de probar que necrosis cardíacas, del tipo de los infartos, pueden ser producidas sin obstrucción vascular, por ejemplo, mediante un tratamiento combinado de corticoides y electrólitos o corticoides y stress; en ese tiempo, con la notable excepción del doctor Sodi-Pallares, pocos de los participantes estaban preparados para creer que tales factores metabólicos pudieran desempeñar un papel en la génesis de verdaderos infartos cardíacos en el hombre. Aunque el efecto funcional de muchos electrólitos sobre el corazón se conoce desde hace tiempo, el hallazgo de que el sodio aumenta y el potasio y el magnesio disminuyen la susceptibilidad a la necrosis cardíaca fue también recibido con la mayor reserva".

Por esos años, en Francia, el Dr. Henri Laborit describía los efectos metabólicos favorables de una solución a base de glucosa, insulina y potasio, a la que llamó "solución polarizante". La insulina favorecía la entrada de la glucosa y del potasio al sarcoplasma (protoplasma) de las fibras miocárdicas, así como la salida del exceso de sodio que había en las mismas fibras en condiciones de isquemia experimental.

Todo parecía concordar: la dieta baja en sodio y alta en potasio descrita por uno de nosotros (Sodi), el efecto agresor del Na^+ y protector del K^+ en las necrosis experimentales de Hans Selye y los efectos favorables en los movimientos iónicos que producía la solución polarizante de Laborit.

El siguiente paso fue el estudio experimental de Sodi-Pallares y colaboradores de la solución polarizante en el infarto reciente experimental producido en el perro por ligadura de la arteria coronaria descendente anterior. A ello nos vamos a referir en seguida.

La solución polarizante

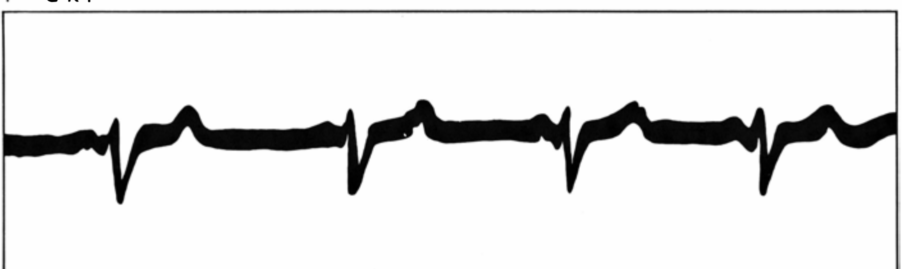
Sodi-Pallares y colaboradores prepararon una solución polarizante un poco diferente de la descrita por Henri Laborit: 1000 ml de solución glucosada al 10 por ciento en los que se disolvían 20 unidades de insulina simple y 40 miliequivalentes de cloruro de potasio (KCl). Esta solución se pasaba por vía venosa a una velocidad de 40 gotas por minuto. El animal del experimento era el perro, al que se le ligaba la arteria coronaria descendente anterior, después de exponer el corazón abriendo el tórax a nivel de la línea medio esternal. Se estudiaron numerosos parámetros comparando corazones con solución y sin solución polarizante, o se esperaba a que las anorma-

I

V1

V6

I G-K-I



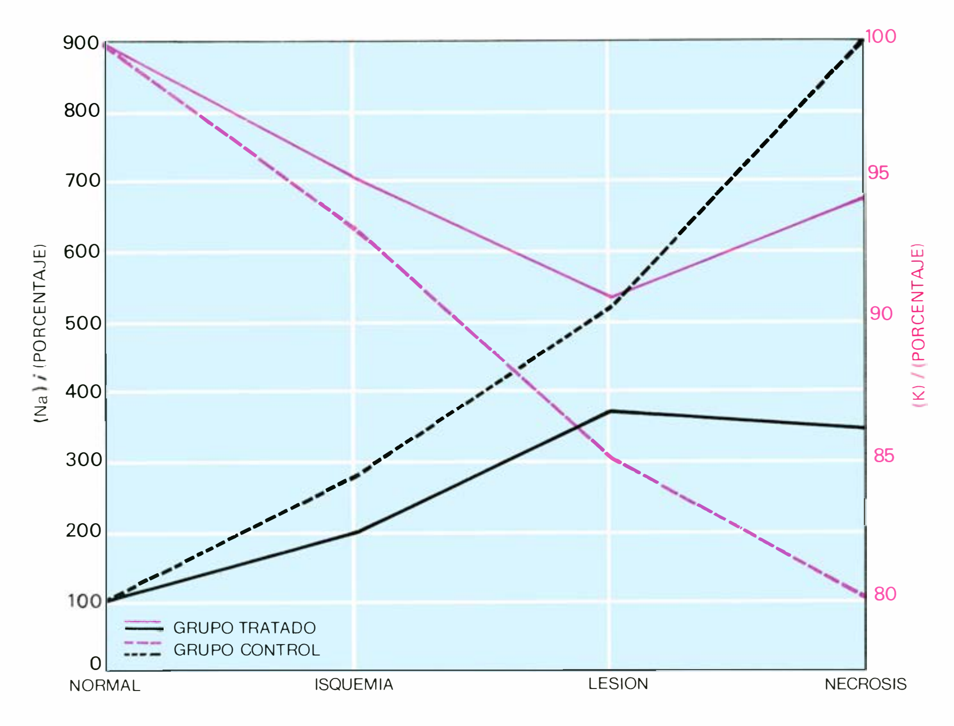
ELECTROCARDIOGRAMA de un paciente con miocarditis (inflamación del miocardio) aguda e infarto del miocardio. Entre cada dos contracciones de la musculatura cardiaca transcurre 1/6 de segundo. La contracción de la pared cardíaca se denomina sístole y diástole su relajación; en cada latido, a la diástole de los ventrículos con sístole simultánea de las aurículas sigue una sístole ventricular acompañada de diástole auricular. Si se aplican estímulos artificiales durante la sucesión normal de latidos se producen contracciones prematuras o extrasístoles. En VI aparece un colgajo de extrasístoles ventriculares. Existe también bloqueo de la rama derecha. En la derivación I *(trazo superior)* hay un desnivel positivo primario de *RS-T* (*RS* indican la propagación de la excitación en el sistema de conducción de estímulos y *T* el retroceso de la misma). Las alteraciones son función del grado de despolarización miocárdica. Con la solución polarizante (GKI –glucosa, potasio e insulina), las extrasístoles y las alteraciones primarias de *RS-T* en DI desaparecieron *(curva inferior)*.

lidades funcionales, que sobre el corazón determinaba la ligadura, aparecieran sin variaciones (constantes experimentales) para estudiar el efecto de la solución. Aquí nos limitaremos a llamar la atención sobre la mejoría de los parámetros más importantes.

(*1*) A las seis horas de la ligadura del vaso coronario, se hace notable la diferencia en el aspecto macroscópico de la superficie epicárdica de un corazón bajo el efecto de la solución polarizante en comparación con otro corazón dejado a la evolución propia del infarto que produce la ligadura. En el corazón sujeto al efecto polarizante se mantiene el color natural del músculo o cambia a un color ligeramente violáceo; en cambio, en la mayoría de los corazones no sujetos al efecto protector de la solución polarizante, la necrosis consecutiva a la ligadura aparece negra en zonas extensas de la superficie epicárdica.

(*2*) Prácticamente todas las arritmias o alteraciones del ritmo cardíaco que aparecen en las primeras horas y días de un infarto reciente del miocardio se controlan fácilmente con la solución polarizante. Una de las condiciones más temibles del infarto agudo es la fibrilación ventricular, situación en que los impulsos atraviesan los ventrículos cardíacos tan rápidamente que la coordinación de las contracciones ocurre raramente. La hipoxia (oxigenación insuficiente) miocárdica baja el umbral de fibrilación a niveles peligrosos, y esta complicación, que evita toda contracción efectiva, aparece cuando dicho umbral ha descendido más allá de un nivel crítico. La solución polarizante eleva el umbral de fibrilación y evita el peligro en la mayoría de las veces. Las extrasístoles ventriculares, aisladas o en colgajos, y las más peligrosas taquicardias ventriculares permanentes, quedan controladas también con la misma solución. Más aún, el peligro de bloqueo auriculoventricular de los infartos posteriores difícilmente se presenta si los pacientes reciben la terapéutica polarizante que, como veremos, mejora el metabolismo celular dañado por el infarto. En nuestra larga experiencia con todo tipo de infartos del corazón sólo por excepción hemos tenido que recurrir a los marcapasos.

(*3*) Tuvimos oportunidad de medir el sodio y el potasio intracelulares a las seis horas de producido un infarto miocárdico experimental. De la periferia al centro del infarto, el aumento de sodio y la disminución del potasio son progresivos. Estas concentraciones iónicas anormales mejoran considerablemente con la solución polarizante de glucosa, insulina y potasio; empero no se normalizan y la necrosis, aunque de menor tamaño, se constituye, como hemos de ver.



EN EL INFARTO RECIENTE del miocardio, el Na^+ intracelular aumenta en tanto que el K^+ intracelular disminuye *(líneas punteadas)*. **Con la solución polarizante, estas alteraciones electrolíticas son menores** *(líneas continuas)*. **El autor, con sus colaboradores, llegó a la conclusión de que las anomalías encontradas en el infarto reciente del miocardio tenían por común denominador una disminución en la energía libre. Por tanto, un tratamiento razonable para este padecimiento debe estar orientado a aumentar la entalpía y la energía libre; sin embargo, un tal enfoque termodinámico no mereció la consideración de la cardiología clásica, a excepción del profesor Hans Selye.**

(*4*) En el tejido infartado la amplitud de la contracción miocárdica disminuye hasta en un 80 o 90 por ciento, e incluso puede desaparecer. Las soluciones polarizantes mejoran la amplitud de la contracción en forma impresionante, como no se ha conseguido con ningún tonicardiaco, incluyendo los digitálicos.

Sodi y colaboradores seguían trabajando contracorriente. El mundo de la cardiología hacía oídos sordos a sus tratamientos. Parecía que sus ideas iban a morir, cuando Eugene Brawnwald demostró que las soluciones polarizantes disminuían el tamaño del infarto en forma impresionante y se protegía la estructura de las miofibrillas (fibrillas musculares) y de los sarcómeros (segmentos en que se divide la fibrilla muscular). Su trabajo llegó a tener tal resonancia que el investigador fue invitado a la Universidad de Harvard a continuar sus estudios y recibió una bolsa de investigación del American College of Cardiology.

Numerosos investigadores han confirmado, hoy en día, los estudios de los médicos mexicanos; empero, el tratamiento polarizante, aunque aceptado como serio por la mayoría de investigadores y cardiólogos, no ocupa aún el lugar preeminente que le corresponde dentro de la terapéutica de las enfermedades del corazón.

¿Por qué creemos que las soluciones polarizantes deben ocupar ese primerísimo lugar en el armamentarium cardiológico? Porque este tratamiento estimula las tres vías metabólicas principales que mantienen una organización celular normal: *a)* aumenta la energía potencial al activar el ciclo del glucógeno, energía que el corazón requiere en momentos de mayor trabajo; *b)* activa la vía de la ribosa y favorece la formación de proteínas, así como de los ácidos ribonucleico y desoxirribonucleico, portadores éstos de los mensajes genéticos y hacedores de las cadenas polipeptídicas que constituyen las nuevas proteínas; *c)* activa la glicolisis anaeróbica y aeróbica (ciclo de Krebs) para formar la molécula en donde se almacena la energía libre (energía cinética de Einstein), es decir, el famoso adenosintrifosfato (ATP) que el corazón requiere constantemente para mantener sus funciones principales incluyendo su propio metabolismo celular.

Lionel Opie, distinguido investigador sudafricano, sugiere diez diferentes posibilidades de acción de las soluciones polarizantes para explicar su efecto benéfico. Ellas son: repolarización celular al aumentar el potasio intracelular; activación de la glicólisis anaeróbica; aumento del glucógeno tisular; aumento del α-gli-

cerofosfato tisular que reesterifica los ácidos grasos libres tisulares; disminución de los ácidos grasos libres circulantes; efectos favorables de "membrana"; efectos hiperosmolares; activación del metabolismo oxidativo de la glucosa; actividad lisosómica disminuida y aceleración de la curación de las heridas. Todas ellas están relacionadas con la activación de las tres vías metabólicas principales.

Ninguna otra terapéutica médica en cardiología, entre las clásicamente aceptadas, tiene esta polifacética acción termodinámica que disminuye la entropía y mantiene la organización de las fibras cardíacas. Intencionadamente señalamos entre las clásicamente aceptadas, puesto que la dieta polarizante, baja en sodio y alta en potasio, es un arma tan poderosa como las soluciones polarizantes.

Dieta baja en sodio y alta en potasio

Usaremos indistintamente dieta baja en sodio y en potasio, dieta polarizante o bien dieta metabólica.

En casos graves, como la hipertensión arterial con cifras diastólicas por encima de 110 mm de Hg, en la angina de pecho, en el infarto reciente del miocardio, en la insuficiencia cardíaca de cualquier grado, en las miocarditis agudas, etcétera, se prescribe la dieta básica que consta solamente de 360 mg de sodio y 2600 mg de potasio. En esta dieta se prohíben los alimentos con más de 100 mg de sodio por 100 g del alimento. Ellos, los prohibidos, son: pan de todas clases, pasteles, galletas con sal, clara de huevo, queso, requesón, natas, crema, mantequilla salada, flanes, gelatinas, jamón, salchichas, chorizo, salami, todos los alimentos embutidos, jaiba, ostión, camarón, sardinas, anchoas, angulas, pulpos, bacalao, todos los mariscos, todos los alimentos enlatados, de frasco, de botella o caja, Corn Flakes, Krispies, All Bran, hot cakes, waffles, agua mineral, cerveza, cacahuetes, nueces, pistaches, "botanas" en general, apio, acelgas, betabel, espinaca, chocolate, bicarbonato de sodio, sales digestivas, pasitas, frutas almibaradas y enlatadas, caldos y consomés. La dieta es rica en frutas, que son los alimentos que aportan el potasio, particularmente la naranja, la toronja, el melón, la uva, el plátano, el higo, la manzana, la fresa y otras; entre las verduras, el tomate es también rico en potasio.

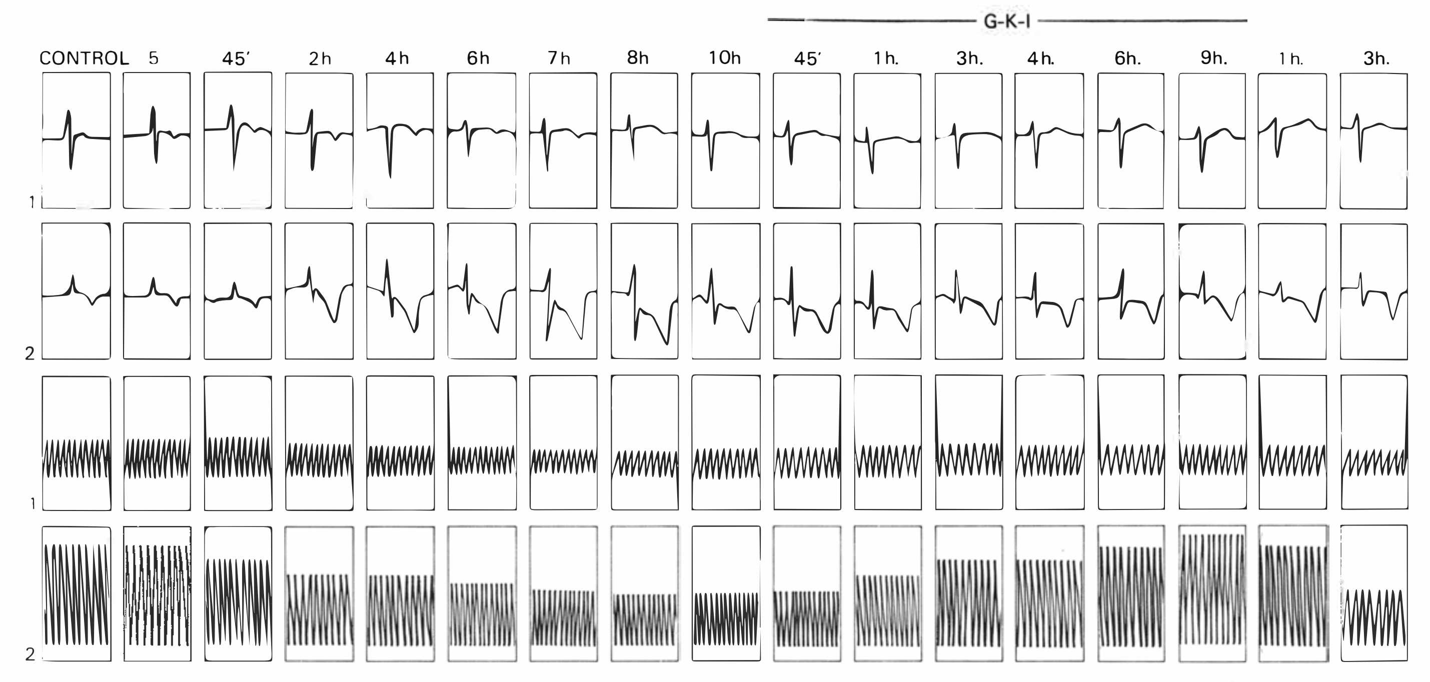
Un menú de la dieta básica podría ser el siguiente: *Desayuno:* un vaso de jugo de naranja; otra fruta, por ejemplo, plátano; un vaso de leche natural de 150 ml; galletas sin sal, a las que se puede untar mantequilla sin sal y mermeladas hechas en casa; una yema de huevo. *Comida principal:* un vaso de jugo de tomate natural; arroz frito con aguacate (en México se agrega salsa picante hecha en casa); 150 g de carne (res, pescado blanco, pollo, pavo, etcétera); una ensalada de verduras (exceptuando las prohibidas) preparada con aceite y vinagre. En México se permiten al paciente alimentos típicos como enchiladas, enfrijoladas, etcétera. Fruta al gusto. De postre, ciruelas pasas españolas, lavadas previamente. *Cena,* como el desayuno.

Esta dieta es rica en grasas (huevo, mantequilla, aceites, etcétera) y, sin embargo, la disminución del colesterol y de los triglicéridos altos en la sangre es la regla. Más aún, hay casos en que tanto la dieta baja en colesterol como el Atromid (clofibrato) se mostraron ineficaces para corregir la hiperlipidemia (exceso de grasas y lípidos en sangre) y, en cambio, la dieta polarizante redujo las cifras de lípidos en forma impresionante. ¿A qué es debido este resultado inesperado? A la estimulación de la vía metabólica que forma la molécula de ATP, cargada de energía libre (negatoentropía), tal y como lo demostró J. W. Gibbs. Vale la pena detenerse en ello brevemente.

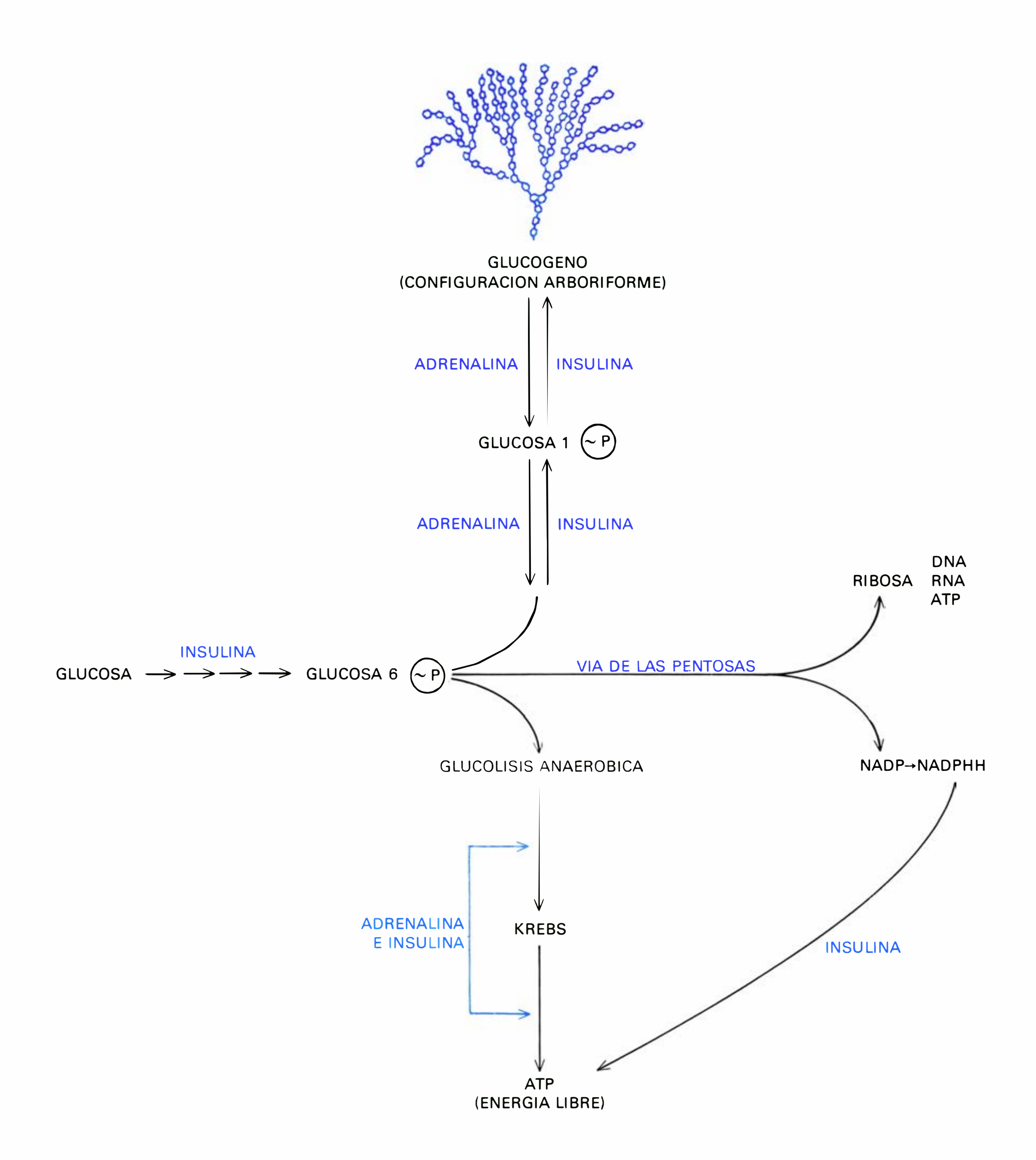
Josiah Williard Gibbs, que sintetizó en una ciencia, la termodinámica, los conocimientos dispersos que sobre energía se tenían, estudió en el siglo pasado los sistemas termodinámicos abiertos, como son los biológicos, y encontró que la energía libre (G) es función de la concentración de los iones sodio $[Na^+]$ y potasio $[K^+]$ en los medios intra y extracelular; función que se expresa como sigue:

$$\Delta G = K \log_e \frac{K^+i / K^+e}{Na^+i / Na^+e}$$

El aumento del potasio intracelular $[K^+i]$ favorece el incremento positivo de la energía libre ($\Delta G\uparrow$); en tanto que el aumento del sodio intracelular $[Na^+i]$ deter-



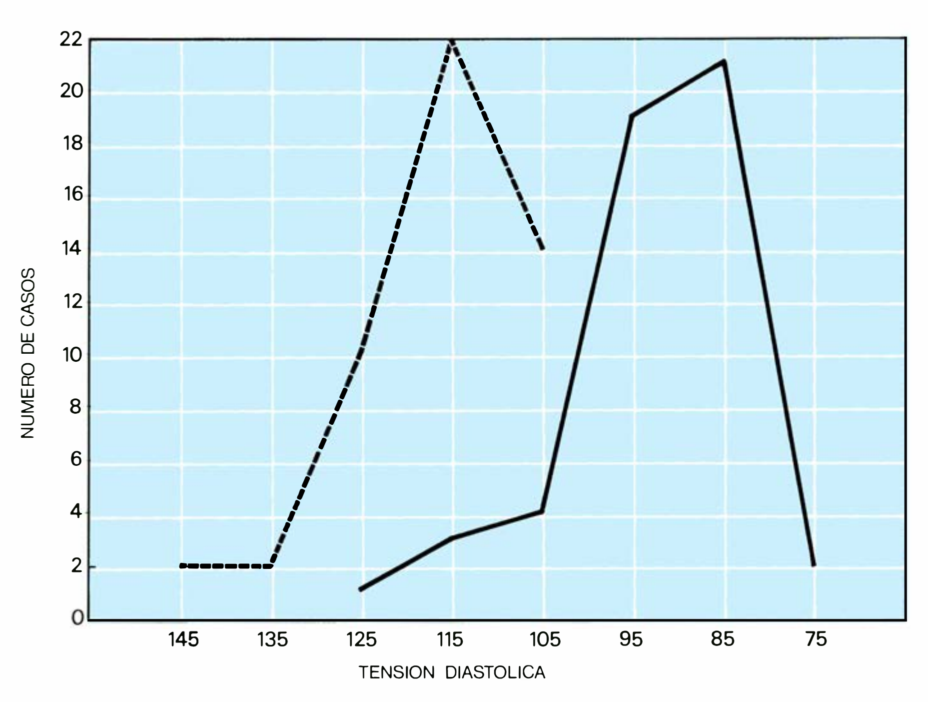
EFECTIVIDAD de las soluciones polarizantes para aumentar la amplitud de la contracción miocárdica que estaba reducida en la zona infartada *(trazos inferiores)*. Cuando las soluciones de glucosa, insulina y potasio se dieron a animales con infarto reciente del miocardio, el potasio intracelular disminuyó en proporción menor, que en otros sin soluciones polarizantes; igual ocurrió con los incrementos de sodio. En consecuencia, las relaciones potasio intracelular/potasio extracelular y sodio intracelular/sodio extracelular, mostraron una mejoría importante. Al suspenderse la solución, la contracción ventricular disminuyó de nuevo, demostrándose así la relación causal entre solución y mejoría.



ESQUEMA TERMODINAMICO de las principales vías metabólicas que nacen a partir del compuesto glucosa-6-fosfato (G-6-P). Puesto que el corazón es un órgano muscular que debe estar contrayéndose casi constantemente, tiene que tener un fuerte almacenamiento de energía potencial, esto es, de glucógeno. Para convertir el glucógeno en ATP o energía libre, primero se rompe la ramificación en unidades de glucosa por acción de un enzima derramificante, interviniendo a su vez otro enzima, la fosforilasa: resultando, pues, glucosa-1-fosfato (G-1-P). Esta pasa a G-6-P. Por la vía de la glicólisis se llega al ácido pirúvico; de éste al ciclo de Krebs o ciclo del ácido cítrico, cuyo rendimiento energético queda plasmado en la figura. A partir de la G-6-P puede seguirse otra vía alternativa: el ciclo de las pentosas. Las vías están esquematizadas.

mina el incremento negativo de la misma energía ($\Delta G\downarrow$). K es la constante de equilibrio de la reacción. Numerosos trabajos experimentales demuestran el efecto activante del K^+ en varios pasos de la vía metabólica que forma el ATP. Bástenos decir que, casi un siglo después de los trabajos originales de Gibbs, muy recientemente (*Fusión*, julio-agosto, 1977), Lawrence Minkoff, al estudiar las interacciones citoplasmáticas de orden elevado, por medio de la resonancia magnética nuclear, afirma que mientras al potasio se le debe ver como "structure maker" (hacedor de estructuras) el sodio es "structure breaker" (rompedor de estructuras).

La dieta metabólica que hemos sugerido reduce la natremia (concentración de Na^+ en el suero sanguíneo) a valores normales bajos, alrededor de 136 miliequivalentes y mantiene la kalemia (concentración del K^+ en el suero sanguíneo) en valores de 4,5 o 5 miliequivalentes. Estas concentraciones estimulan los mecanismos de la bomba biológica que extrae el sodio intracelular y favorece el paso del potasio al interior de la célula; o lo que es lo mismo, es una dieta que estimula, de acuerdo con la fórmula de Gibbs, la formación de energía libre y de la vía metabólica principal, la que forma el ATP, molécula vital negato-



CIFRAS DE TENSION DIASTOLICA obtenidas antes *(línea punteada)* **y después** *(línea continua)* **del tratamiento con dieta estricta hiposódica (baja concentración en sodio) e hiperpotásica (alta concentración en potasio), en 50 enfermos con hipertensión arterial. Las edades de los 28 hombres y 22 mujeres del grupo oscilaron entre 42 y 75 años, con edad promedio de 56 años.**

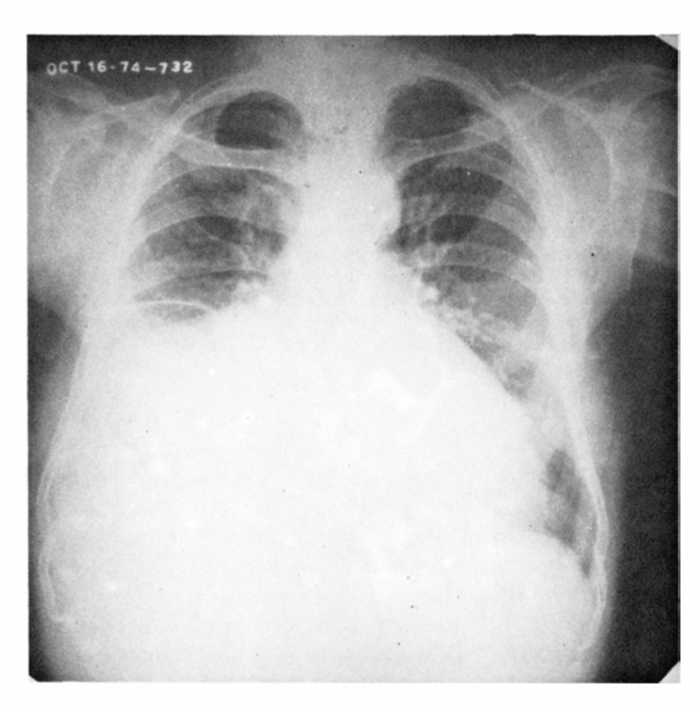
entrópica. Por esto, a nuestra dieta baja en sodio y alta en potasio la hemos llamado: dieta metabólica o dieta polarizante.

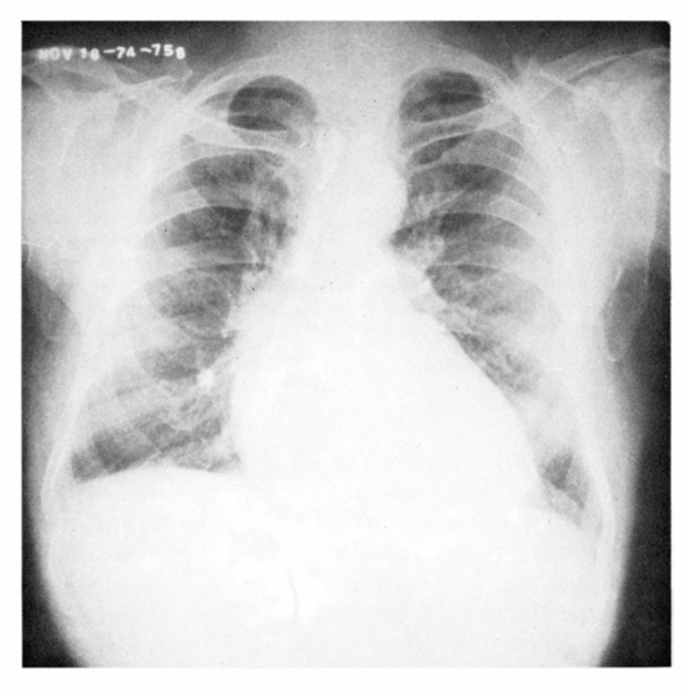
Los resultados que hemos obtenido con esta dieta que uno de nosotros (Sodi) empezó a prescribir a sus enfermos hace 33 años, son sorprendentes; particularmente en la hipertensión arterial, con caída rápida (en 1 o 2 semanas) de las presiones diastólicas; en la angina de pecho, con desaparición del dolor en pocos días, rara vez se requieren pocas semanas; en las insuficiencias cardíacas, sin necesidad de recurrir a digitálicos y diuréticos, a menos que se trate de casos muy avanzados o con complicaciones, en las cuales asociamos nuestra dieta metabólica con las soluciones polarizantes.

A modo de ejemplo, permítasenos aducir un caso que ilustra lo que se puede conseguir si se estimula el metabolismo celular en forma adecuada. Con autorización del paciente, le llamaremos el caso del "señor Gallardo". Hoy en día, el paciente tiene 77 años de edad. Hace nueve años, lo tratamos de insuficiencia cardíaca grave con dieta y soluciones polarizantes. Una vez controlada la insuficiencia, el enfermo quedó con la dieta baja en sodio y alta en potasio como único tratamiento. El enfermo estuvo bien mientras siguió la dieta; pero la abandonó en 1974 y en pocos meses reapareció la insuficiencia cardíaca, más grave que la primera vez, pues en esta ocasión había anasarca: ascitis y derrame pleural. De nuevo el tratamiento polarizante (dieta y soluciones) hizo desaparecer todo signo de insuficiencia cardíaca, incluyendo el derrame pleural. Lo notable del caso sucedió posteriormente. El señor Gallardo, a partir de la segunda mejoría, siguió estrictamente la dieta y empezó a realizar ejercicios, con entrenamiento adecuado; es decir, aumentaba el ejercicio, a pesar del bloqueo avanzado de rama derecha y de la fibrilación auricular que mostraban sus electrocardiogramas. Ahora, a los 77 años, el señor Gallardo es campeón de la marcha acelerada de los 3 kilómetros (competencia que se lleva a cabo en el Bosque de Chapultepec de México).

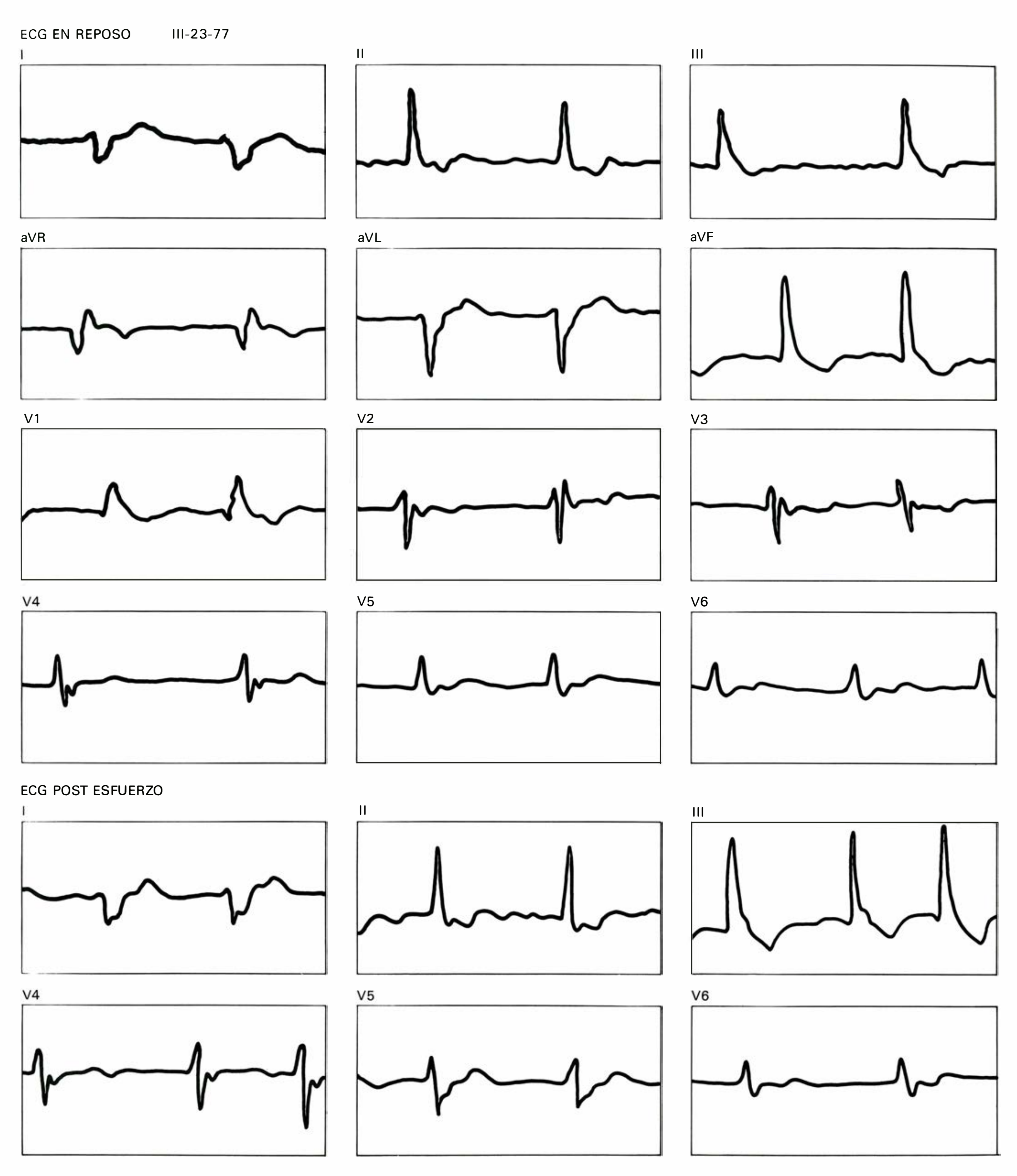
Otros aspectos del tratamiento metabólico

Como complementarios de la dieta y solución polarizante, para mejorar o





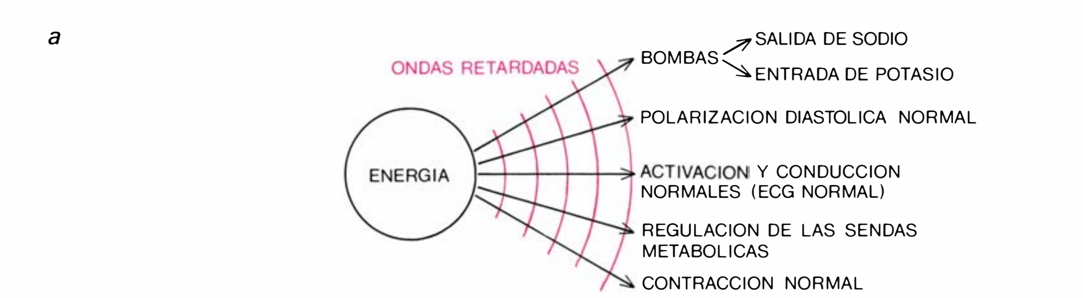
CASO CLINICO estudiado por el autor, de extraordinaria relevancia. A la izquierda puede observarse un aumento importante en el tamaño del corazón (cardiomegalia grado 3) y el considerable derrame pleural derecho; ocurría así antes del tratamiento. A la derecha, puede observarse la notable mejoría, tanto en la desaparición del derrame pleural como en la reducción de la cardiomegalia, o hipertrofia del corazón.

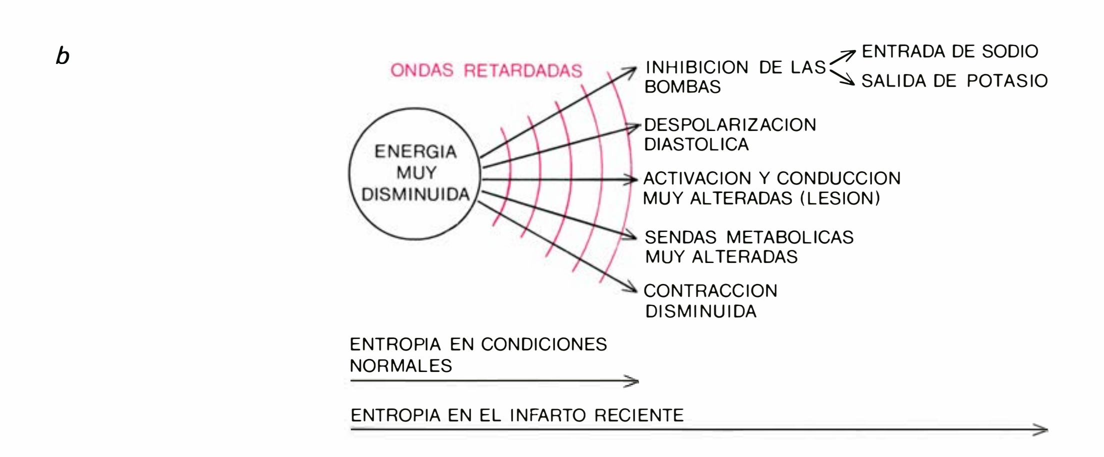


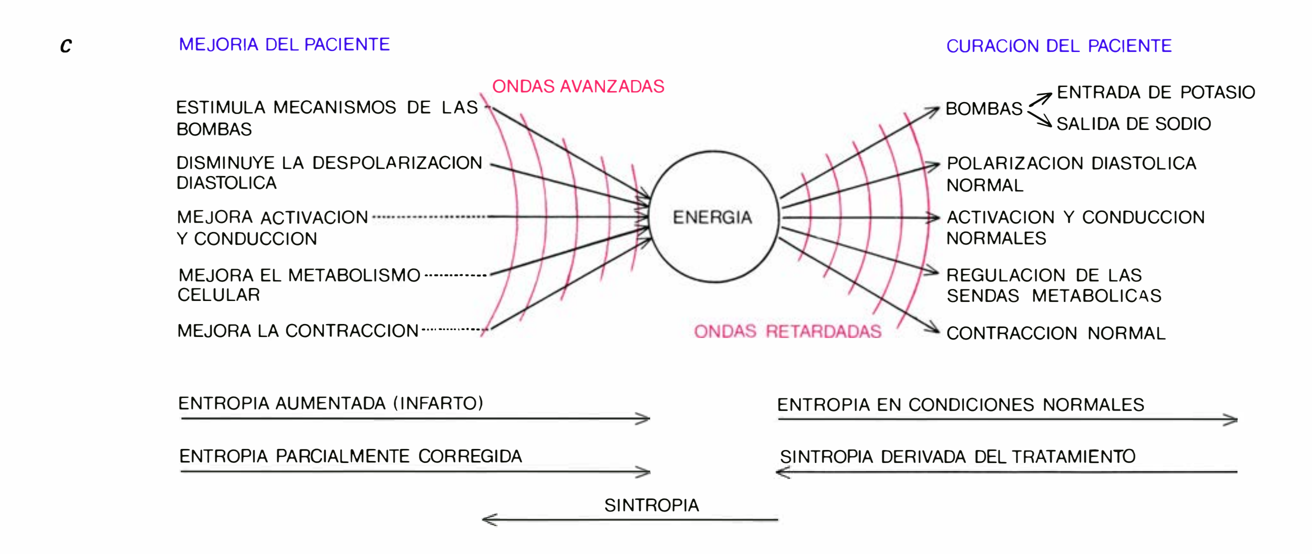
ELECTROCARDIOGRAMA que muestra un bloqueo avanzado de la rama derecha del haz de His (banda muscular que contiene fibras nerviosas que conexionan las aurículas con los ventrículos cardíacos y sirve para transportar los estímulos de aquéllas a éstos) y fibrilación auricular que está controlada sin digital, que es una planta medicinal vasoconstrictora y diurética. Fibrilación es sinónimo de contracción incoordinada.

normalizar las tres vías metabólicas principales afectadas por la cardiopatía, contamos con los recursos siguientes:

En primer lugar, insulina lenta, inyectada diariamente por vía subcutánea en dosis que oscilan de 3 a 10 unidades. Nos referimos a enfermos no diabéticos, pues si lo son, las dosis son mayores de acuerdo con cada caso. Esta hormona pancreática activa las tres vías metabólicas. Prescribimos la insulina de acción lenta en enfermos con angina de pecho y en las insuficiencias cardíacas graves que no se controlan exclusivamente con la dieta. Para completar la acción de la insulina sobre el aporte energético (paso de glucosa, aminoácidos y grasas a través de la membrana de las fibras cardíacas), se aconseja al paciente incrementar los azúcares (miel o mermeladas hechas en casa) en el desayuno, que irá seguido de la in-







MODELOS TERMODINAMICOS que explican la variabilidad de los distintos parámetros del corazón en función de la energía libre obtenida por hidrólisis del adenosintrifosfato (ATP): en condiciones normales *(a)*, en infarto reciente del miocardio *(b)* y al concluir el tratamiento polarizante *(c)*. En condiciones normales *(a)* existe un ciclo de retroalimentación regulador en el cual la energía se aprovecha para los mecanismos de bomba. En efecto, la membrana celular puede atravesarse por los cationes Na^+ y K^+, para cuyo desplazamiento requieren energía; si bloqueamos el suministro energético, la concentración de potasio del interior de la célula disminuye y aumenta la concentración de sodio, hasta que se igualan las concentraciones en el interior y en el exterior de la célula; al restablecer el suministro de energía, la membrana reanuda su actividad en el sentido de que concentra potasio en el interior de la célula y transporta sodio fuera de ella; el proceso se denomina transporte activo y, debido a que la membrana bombea iones contra un gradiente existente, se dice que se comporta como una bomba. La polarización, o potencial de reposo, depende de la eficacia de los mecanismos de bomba; bajo condiciones de equilibrio termodinámico, los iones deben mantenerse a concentraciones diferentes a ambos lados de la membrana, equilibrio que implica la generación de un potencial, dada la diferencia entre el grado de difusión. La activación y la conducción depende de la polarización normal (si el potencial de reposo es menor que el normal, aumenta el automatismo para luego deprimirse, y la conducción se retarda). Las vías metabólicas requieren la aportación energética: la glucosa necesita ATP para atravesar la membrana. El proceso se altera en el infarto reciente del miocardio *(b)*, en donde se advierte un incremento exagerado de entropía, para regularizarse una vez se da por finalizado el tratamiento polarizante *(c)*.

gestión de cloruro de potasio en comprimidos. Podríamos decir que se trata de una combinación de glucosa-insulina-potasio menos enérgica que la de las soluciones polarizantes; en cambio, está más al alcance del cardiólogo y del internista en su práctica diaria.

En segundo lugar, propanolol (Inderal) beta-bloqueador; protege al corazón de la acción catecolamínica agresora. Las catecolaminas son las hormonas de la médula que comprenden la adrenalina y la noradrenalina; ambas favorecen el cambio de energía potencial a energía libre, cada vez que el corazón es requerido por el organismo para realizar mayor trabajo (ejercicio, stress, fiebre, etc..). Sin embargo, la acción adrenalínica en exceso es metabólico-agresora, lo que se reconoce fácilmente en el electrocardiograma por taquicardia, extrasístoles auriculares y/o ventriculares, y por desniveles negativos de RS-T, sobre todo, en las derivaciones ventriculares izquierdas. En estos casos, el propanolol, prescrito en la dosis adecuada, evita la acción agresora catecolamínica, sin cortar el efecto adrenalínico normal que el corazón requiere en las emergencias. El propanolol es, pues, protector metabólico y se ha demostrado que incrementa la producción de ATP cuando su indicación es precisa.

En tercer lugar, hormonas tiroideas: solamente en casos de función glandular baja. La tiroxina es necesaria para el corazón; en particular, en la vía que forma proteínas y ribosa, por lo que también favorece la formación de la molécula fundamental que almacena energía libre es decir, el ATP. La combinación de tiroxina con propanolol es de empleo frecuente en el tratamiento metabólico de las cardiopatías de nuestros pacientes.

Por último, ejercicio metódico y entrenamiento adecuado, que uno de nosotros (Sodi) ha diseñado para las diferentes partes del cuerpo: dedos, muñecas, brazos, hombros, cuello, cara, abdomen piernas y sectores diferentes de la columna vertebral.

Recapitulación

Resumiendo, el tratamiento metabólico (termodinámico) que prescribimos en las cardiopatías principales consta de los apartados siguientes: Soluciones polarizantes, dieta baja en sodio y alta en potasio, insulina de acción lenta, propanolol si hay efecto catecolamínico agresor; hormona tiroidea si hay hipofunción de la glándula y ejercicio metódico de las diferentes partes del organismo.

Estrellas de rayos X en cúmulos globulares

Los densos núcleos centrales de algunos cúmulos favorecen la formación de estrellas dobles emisoras de rayos X, en las cuales una estrella de neutrones o un agujero negro capturan materia de otra estrella que está consumiendo su combustible nuclear

George W. Clark

Durante los últimos 15 años, los telescopios de rayos X transportados por globos, cohetes y, más recientemente, por observatorios instalados en satélites, han localizado unas 100 estrellas de rayos X, altamente luminosas y variables, entre los 100.000 millones de estrellas que forman el disco de nuestra galaxia. Una típica estrella de rayos X radia, en esa región del espectro, con una potencia que es miles de veces superior a la que radia el Sol en todas las longitudes de onda. Con una frecuencia de aproximadamente uno por cada mil millones de estrellas, tales objetos se cuentan entre los más raros y fascinantes de todos los fenómenos celestes. Como resultado de las observaciones relativamente precisas, suministradas por telescopios de rayos X a bordo de satélites, se ha establecido en la actualidad que seis estrellas de rayos X por lo menos, variables y persistentes, están situadas dentro de cúmulos globulares de estrellas, en la galaxia.

Los cúmulos globulares son sistemas autogravitantes que constan, por término medio, de unas 100.000 estrellas. De ellos se conocen, en la actualidad, unos 130, y cabe que otros 70 más se hallen velados por nubes de polvo interpuestas entre ellos y nosotros. Pueblan el " halo" galáctico, que es la región esférica bisecada por el disco de la galaxia. En total, los cúmulos globulares constan de unos 10 millones de estrellas. Por tanto, la presencia en ellos de seis estrellas de rayos X representa una frecuencia de aparición de aproximadamente una cada dos millones, proporción 500 veces mayor que la de las estrellas de rayos X en el disco de la galaxia. La mayor frecuencia debe indicar un ritmo de formación de estrellas de rayos X más elevado, en los cúmulos globulares, mayor longevidad de esos objetos, o ambas cosas.

Más recientemente, los telescopios de rayos X instalados a bordo de satélites han descubierto más de 30 "eruptores" de rayos X; los "eruptores" constituyen un tipo especial de estrella de rayos X que emite breves y brillantes impulsos de rayos X, cuya duración alcanza unos cuantos segundos. En algunos casos, los impulsos se repiten a intervalos bastante regulares, que se miden por horas o por días. Hay un solo "eruptor" rápido que emite sus impulsos a fuego graneado, a ritmos que llegan a varios miles por día. En una erupción típica de 10 segundos, el flujo de rayos X puede sobrepasar el equivalente a una semana del flujo total de energía del Sol a todas las longitudes de onda. Se han detectado emisiones persistentes, altamente variables, de rayos X procedentes de varios de los "eruptores" lentos, lo cual demuestra que las erupciones representan una forma extrema de variación, característica de un tipo particular de estrella de rayos X.

La mayoría de los "eruptores" están situados dentro del disco galáctico y se acumulan hacia el centro galáctico. Se han encontrado, sin embargo, tres "eruptores" lentos en cúmulos globulares de los que ya se sabía que contenían estrellas de rayos X persistentes. El único "eruptor" rápido está en la dirección de un cúmulo globular desconocido anteriormente, que queda casi oculto por el polvo interestelar. Así pues, el "eruptor" rápido eleva a siete el número de estrellas de rayos X en cúmulos globulares. Por lo menos en uno de los casos de "eruptores" lentos situados en cúmulos globulares, las erupciones se observan sólo cuando la emisión de rayos X más persistente es bastante débil, lo cual indica que el eruptor y la fuente persistente constituyen el mismo objeto.

Las observaciones realizadas hasta el presente, sobre estrellas de rayos X en cúmulos globulares, no son todavía lo bastante extensas como para poder establecer si todas ellas emiten ocasionalmente por erupciones, o no lo hacen. Por otra parte, varias de las fuentes de rayos X del disco galáctico se han observado durante bastantes semanas (en tiempo acumulado), por uno o más de los observatorios en satélites, sin que haya ninguna prueba de erupciones. Aparece, pues, como seguro que algunas, o tal vez la mayoría de las estrellas de rayos X persistentes que no están dentro de cúmulos, tampoco son eruptivas. Resulta claro, por consiguiente, que las especiales circunstancias que dan lugar a estrellas de rayos X en cúmulos globulares favorecen, al mismo tiempo, la producción del tipo que origina erupciones.

¿Qué clase de objeto puede generar rayos X al prodigioso nivel de potencia de las estrellas de rayos X? Podemos afirmar con bastante seguridad que las estrellas de rayos X son residuos densos de estrellas que han agotado sus fuentes de energía nuclear y que han colapsado bajo la fuerza atractiva de su propia gravedad. Su emisión de rayos X proviene de un proceso de captura, en el cual la energía potencial gravitatoria de materia que cae hacia la estrella colapsada se convierte en calor a temperaturas de muchos millones de grados y se radia como fotones de rayos X, los cuales tienen energías comprendidas, en su mayor parte, entre 1000 y 50.000 electrón-voltios (entre uno y 50 KeV). Por tanto, para constituir una estrella de rayos X, la naturaleza ha de combinar un residuo estelar denso y un aporte de materia que alimente el proceso de captura a un ritmo adecuado.

Ambos constituyentes vienen proporcionados por un sistema binario que

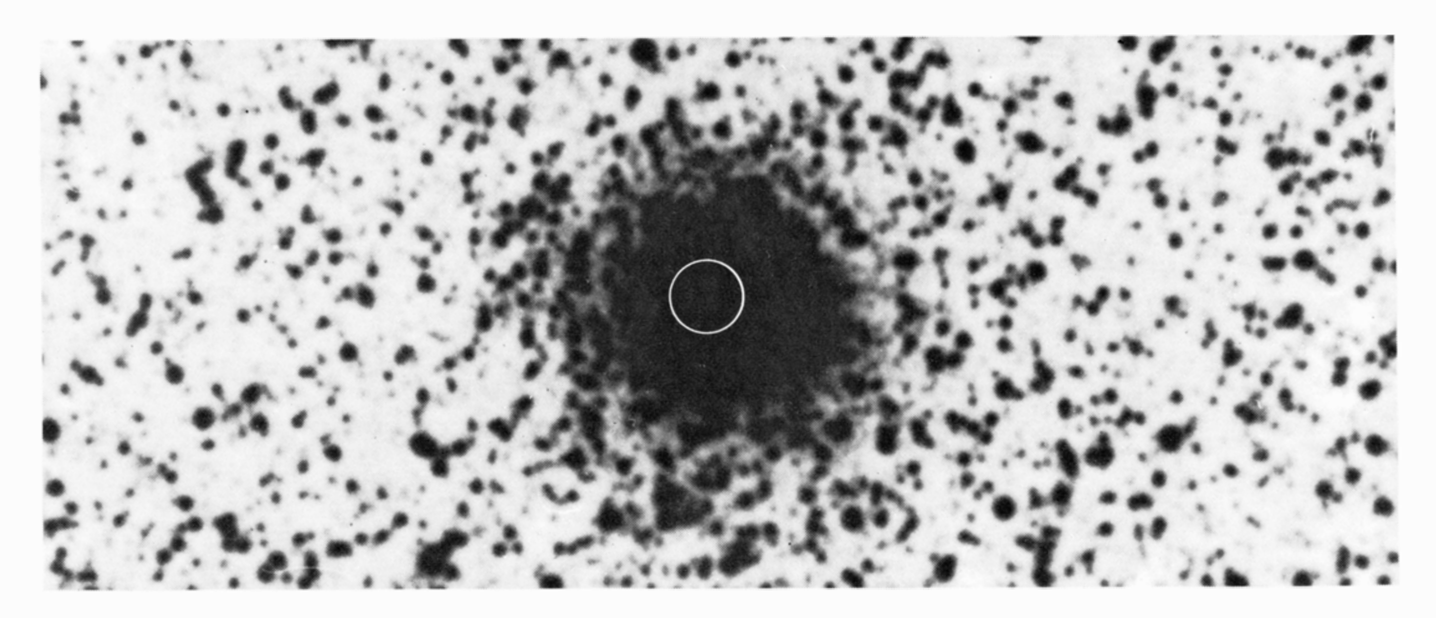
consiste en una estrella que aún está quemando su combustible nuclear, en la inmediata vecindad de una estrella quemada y colapsada. En determinadas condiciones, fluirá materia desde la estrella que quema combustible nuclear hasta la estrella compacta, lo cual mantendrá el proceso de captura. Pueden provocarse erupciones de rayos X a través de interrupciones periódicas del flujo de materia, que podría amontonarse y, luego, de golpe, verterse sobre la superficie de la compañera colapsada. Los miembros densos de esas parejas son estrellas de neutrones, cuya masa equivale aproximadamente a la del Sol y sólo poseen unos 10 kilómetros de radio, o bien son agujeros negros, objetos que están contenidos por entero dentro de un radio crítico, para el cual el campo gravitatorio resulta tan intenso que ni la misma luz puede escaparse de él. El radio crítico para un cuerpo de una masa solar es de unos tres kilómetros, y proporcionalmente mayor para objetos dotados de mayor masa.

Aunque la existencia de los agujeros negros sigue siendo un tema controvertido, no parece que haya ninguna otra vía para que un cuerpo de masa suficiente y de momento angular lo bastante bajo pueda escapar, en último término, a un colapso hacia agujero negro. Como veremos, hay pruebas claras de que algunas estrellas de rayos X son estrellas de neutrones en rotación, con poderosos campos magnéticos que controlan el flujo del material que se captura, produciendo así pulsaciones extremadamente regulares y reveladoras en su emisión de rayos X. Algunas estrellas de rayos X no presentan pulsaciones regulares en absoluto, y pueden ser estrellas de neutrones con campos magnéticos débiles o despreciables o bien agujeros negros. Tal es el caso de las estrellas de rayos X de los cúmulos globulares. De aquí que el estudio de tales objetos suministre información, tanto acerca de la naturaleza de las fuentes no pulsantes de rayos X como sobre la evolución de los cúmulos en los cuales éstas se crean más abundantemente.

Los cúmulos globulares son los objetos más antiguos y maravillosos de la galaxia. Han tenido un papel crucial a la hora de definir el tamaño y la forma de la galaxia como un todo. El primero en reconocerlos como sistemas gravitatoriamente ligados fue William Herschel, en la década de 1780. Un siglo más tarde, Solon I. Bailey descubrió multitud de estrellas en las regiones externas de varios cúmulos globulares cercanos, cuyo brillo variaba periódicamente. Las variables con períodos regulares del orden de un día constituyen la clase de variables RR Cyrae, muy conocida. Dado que todas tienen casi la misma luminosidad media, resultan útiles como bujías convencionales para mediciones de distancias. Usando las variables RR Cyrae de los cúmulos globulares como índices de distancia, Harlow Shapley concluyó, en 1917, que los cúmulos se hallaban distribuidos, más o menos esféricamente, en torno a un punto de la Vía Láctea situado en la dirección de la constelación Sagitario. Shapley propuso que el centro del halo de cúmulos globulares era también el centro de todo el sistema galáctico.

Los trabajos de Shapley constituyeron la base de la opinión actual que sostiene que los cúmulos se formaron hace 10.000 millones a 13.000 millones de años, en el colapso gravitatorio de la protogalaxia, una extensa nube de gas que constaba de hidrógeno y helio. Al contraerse la nube protogaláctica, las regiones de mayor densidad se hicieron autogravitantes y se condensaron, en un tiempo relativamente corto, en cúmulos globulares de estrellas, conteniendo cada uno millones de ellas. Los cúmulos, una vez formados, empezaron a moverse, a través de la nube protogaláctica, como cuerpos cohesionados en órbitas determinadas por el campo gravitatorio general de la galaxia y con períodos orbitales de centenares de millones de años. Al ir ganando velocidad en su movimiento orbital, quedaron despojados de toda la materia que no había condensado ya en estrellas. Por tanto, no aparecieron nuevas estrellas después de la erupción inicial de formación estelar; cabe que el intervalo de edades de las estrellas de un cúmulo globular dado sea de unas decenas de millones de años, es decir, menos del 1 por ciento de su edad total. El gas remanente de la nube protogaláctica siguió contrayéndose, y, en un tiempo comparable con el período orbital de un cúmulo, colapsó en un disco delgado en rotación, dentro del cual se formaron las estrellas de la Vía Láctea. Los cúmulos siguieron moviéndose en amplias órbitas que llenan la región en la cual nacieron, esto es, en el halo esférico de la galaxia.

Se ha perdido gran parte de la masa inicial de los cúmulos globulares. Ya du-



LA MAS BRILLANTE DE LAS SIETE ESTRELLAS de rayos X, descubiertas en siete cúmulos globulares distintos, cae dentro del circulo de error de un radio de 20 segundos de arco, inscrito en este negativo de una fotografía tomada con el telescopio Schmidt de 1,22 metros de Monte Palomar. El cúmulo es el NGC 6624; la estrella de rayos X que contiene es 3U 1820-30. Cada uno de estos cúmulos consta de unas 100.000 estrellas. La posición de 3U 1820-30 se determinó con el satélite SAS-3. El centro del cúmulo está atiborrado de estrellas.

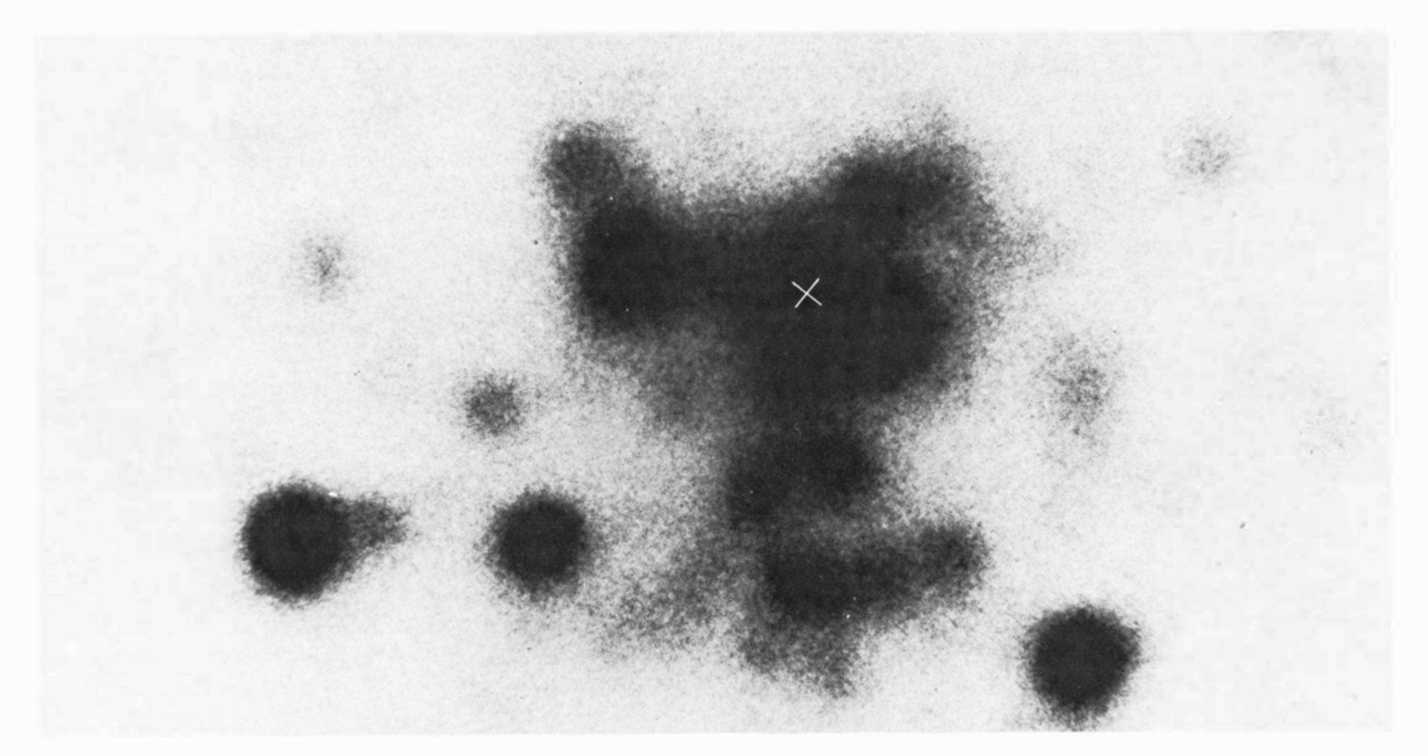
rante el primer período orbital galáctico, cada estrella cuya masa fuera superior a unas cuantas masas solares completó su ciclo de vida y estalló como una supernova que arrojó fuera los elementos pesados que se habían sintetizado en su núcleo por reacciones termonucleares. Este material contaminó el hidrógeno y el helio primordiales de la nube protogaláctica en contracción con una concentración creciente de elementos más pesados; ello trajo por consecuencia que los cúmulos formados luego, que se encuentran ahora más cerca del centro de la galaxia, se hallen enriquecidos en esos elementos.

Las estrellas dotadas de masa menor que poblaban inicialmente los cúmulos originales, estrellas cuya masa no llegaba por poco a ser la del Sol, completaron su ciclo de vida sin estallar, evolucionando finalmente hacia estrellas enanas blancas. Las estrellas de masas muy pequeñas y cuerpos del tamaño del planeta Júpiter o menores fueron expulsados de los cúmulos por fuerzas de marea, en sus repetidos pasos a través del disco; o bien se vieron eyectados por un proceso parecido al de la evaporación, en el cual las estrellas adquirían la velocidad de escape a través de interacciones gravitatorias, al azar, con otras estrellas del cúmulo. Hay razones para creer que una fracción substancial de la masa total de la galaxia está en forma de tales estrellas de masas pequeñas, de débil o nula luminosidad, que antaño estuvieron en cúmulos globulares. Lo que queda, en un cúmulo típico, es una masa total que equivale a la masa de 100.000 soles, aproximadamente; en su mayor parte persiste en forma de estrellas de 0,4 a 0,8 masas solares, las cuales están quemando aún su combustible nuclear. El resto del cúmulo consta de residuos mortecinos, o bien ópticamente no luminosos, de las estrellas dotadas de mayor masa que consumieron su combustible nuclear y colapsaron. Las observaciones en rayos X han comenzado ahora a suministrar información directa sobre algunos de estos miembros densos y ópticamente obscuros, de los cúmulos globulares.

Examinemos ahora con más detalle los procesos de formación y de evolución estelares, a fin de ver cómo pueden surgir estrellas de rayos X, tanto dentro del disco galáctico como dentro de los cúmulos globulares. La formación de estrellas empieza con la contracción de una nube protoestelar de gas y polvo. En el disco galáctico la mitad aproximada de las estrellas han nacido de una en una. El resto lo ha hecho en sistemas dobles o múltiples. En la mayoría de los casos, los miembros de tales sistemas están lo bastante separados como para que evolucionen igual que las estrellas aisladas, sin verse afectadas por sus compañeras. Algunas estrellas binarias están, sin embargo, tan juntas que la evolución de ambas se ve drásticamente alterada por la transferencia de materia. Bajo ciertas condiciones, tales sistemas binarios pueden convertirse en poderosas fuentes de rayos X. Primero, no obstante, trazaremos la evolución de las estrellas nacidas de una en una.

Una nube protoestelar se contrae a un ritmo que está regulado inicialmente por la transformación de la energía gravitatoria en calor. Con el tiempo, la temperatura y la presión en el núcleo de la nube en contracción llegan a ser lo bastante altas para iniciar las reacciones de fusión que convierten los núcleos de hidrógeno en núcleos de helio. La liberación de energía nuclear que resulta de ello detiene eficazmente la contracción de la estrella y es causa de que la temperatura y la densidad se mantengan en un estado casi estacionario; dicho estado se caracteriza por una combinación particular de la temperatura superficial y de la luminosidad total, combinación que viene determinada básicamente por la masa total de la estrella.

Una gráfica que represente la temperatura superficial en función de la luminosidad, para multitud de estrellas de diferentes masas que estén quemando hidrógeno en sus núcleos, las mostrará agrupadas a lo largo de una línea conocida por secuencia principal del diagrama

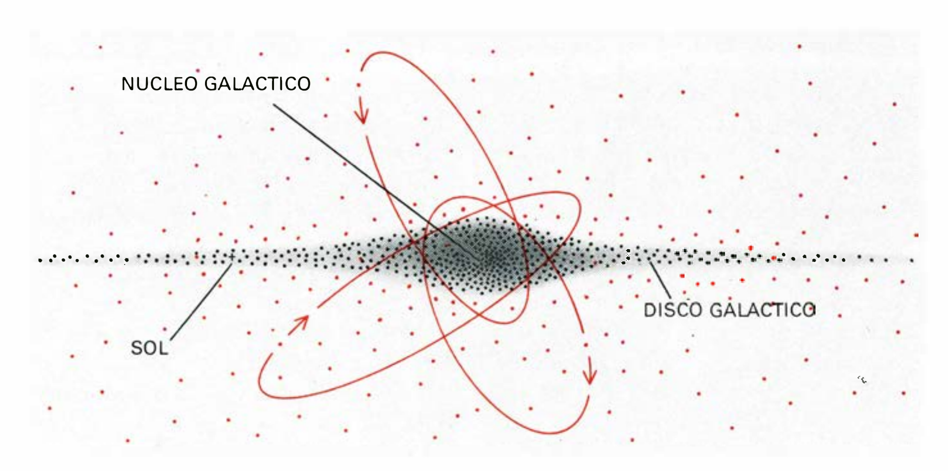


EL NUCLEO CENTRAL del cúmulo globular NGC 6624 aparece en esta fotografía de alta resolución y breve exposición, tomada por William Liller, del Harvard College Observatory, con el reflector de cuatro metros del Observatorio Interamericano de Cerro Tololo, en Chile. Una docena de estrellas supergigantes rojas están apiñadas en una región de sólo 10 segundos de arco de diámetro. A la distancia que se ha estimado para el cúmulo, de 20.000 años-luz, ese diámetro angular corresponde a un año-luz. Por cada supergigante roja hay, probablemente, unas 200 estrellas, que resultan demasiado débiles para quedar registradas en la fotografía. Así pues, la densidad estimada de estrellas en el núcleo del cúmulo es tal que su masa por año-luz cúbico podría equivaler a 3000 veces la masa del sol. Y siguiendo con las comparaciones, tal densidad de estrellas resultaría equivalente a un millón de veces la densidad de las estrellas que se hallan en el entorno solar.

de Hertzsprung-Russell, o diagrama H-R. Cuanto mayor masa posea una estrella, tanto más deprisa evolucionará y mayores serán la luminosidad y la temperatura superficial.

Durante la fase de combustión en el núcleo, la posición de una estrella en el diagrama H-R se desplaza sólo una pequeña distancia hacia arriba, a lo largo de la secuencia principal, de modo que la correspondencia entre masa inicial y posición a lo largo de la secuencia principal depende sólo ligeramente de la edad. Cuando se ha consumido todo el hidrógeno del núcleo de la estrella, ésta entra en un período de complejos reajustes en su temperatura y en su densidad y empieza a quemar el hidrógeno en una capa en torno al núcleo. Durante esta fase, el helio sigue acumulándose en el núcleo, el ritmo de producción de energía aumenta, la envoltura externa se expande, la temperatura de la superficie decrece y la estrella se transforma en una gigante roja. Al representar una estrella de este tipo en el diagrama H-R, se observa que se aparta de la secuencia principal y se mueve rápidamente hacia arriba y hacia la derecha, a lo largo de la rama gigante roja.

Tras haber invertido quizás un 90 por ciento de su vida de quemadora de combustible nuclear, manteniéndose tranquilamente en la secuencia principal, la estrella sufre luego cambios drásticos en los procesos por medio de los cuales genera energía nuclear, culminados por violentas disrupciones de su estructura. Finalmente, cuanto queda es una nebulosa en expansión de gas eyectado y un residuo denso de productos de fusión que han agotado su energía nuclear disponible. Desprovisto de fuente interna de energía, el residuo pierde la presión de la agitación térmica y de la radiación y se hunde en un estado de densidad extremadamente alta.



VISION ESQUEMATICA DE NUESTRA GALAXIA. Muestra cómo los cúmulos globulares de estrellas *(color)* **pueblan un "halo" esférico en torno al disco central de la galaxia** *(vista aquí de canto).* **Un cúmulo globular típico completa una órbita en torno al centro de la galaxia en unos 300 millones de años. A cada paso a través del disco, se barre el gas y el polvo advenidos al cúmulo. El diámetro del disco es de 100.000 años-luz. El sol está a 27.000 años-luz del centro del disco.**

El primer ejemplo de residuo estelar denso fue descubierto en 1862 por Alvan Clark, el mejor constructor de telescopios de su época. Durante una prueba en fábrica de una nueva lente de 47 centímetros, la mayor hasta entonces, captó fugazmente una tenue estrella situada junto a Sirio, la estrella más brillante del cielo, mientras la propia Sirio quedaba velada por el perfil de un edificio lejano. Aquel objeto tenue, que poseía sólo una diezmilésima del brillo de Sirio, resultó ser una compañera binaria de ésta, cuya existencia había sido inferida por Friedrich Bessel en 1844, a partir de mediciones de las perturbaciones del movimiento de la estrella brillante. El análisis de las órbitas mostró que la compañera, Sirio *B*, poseía casi la masa del Sol, pese a que su brillo fuera sólo un 0,2 por ciento del de éste. Cuando Walter S. Adams mostró en 1914 que el espectro de Sirio *B* era el de una estrella blanca, caliente, quedó claro que la estrella debía su carácter tenue a su pequeñez; era aproximadamente del tamaño de la tierra. Por tanto, la densidad de Sirio *B* supera los 100 kilómetros por centímetro cúbico.

R. H. Fowler explicó la estructura de Sirio *B* y otras enanas blancas en 1926, basándose en la mecánica cuántica, entonces recientemente inventada. Su teoría invocaba el principio de exclusión de Pauli, que limita la densidad a la cual pueden empaquetarse electrones cuyos momentos sean inferiores a un valor determinado en un volumen dado del espacio. Se dice que, a esa densidad límite, los electrones se encuentran en un estado de degeneración. Fowler sugirió que las enanas blancas se comprimen por la gravedad hasta la densidad límite para los electrones, y que tales estrellas están efectivamente sostenidas frente a una nueva contracción por la presión de sus electrones degenerados. La teoría detallada, publicada por Subrahmanyan Chandrasekhar en 1931, mostraba que esta fuente de soporte es estable hasta una masa crítica de 1,4 masas solares, para una enana degenerada compuesta de helio o carbono. Debido a su pequeña superficie, tales estrellas pueden seguir brillando como enanas blancas, tenues y visibles, durante miles de millones de años, en virtud de su calor almacenado.

Si un núcleo degenerado excede el límite de Chandrasekhar, colapsará. Lo que sucederá a partir de entonces fue sugerido en 1933 por Walter Baade y Fritz Zwicky, quienes, por entonces, se hallaban buscando una explicación de las supernovas; éstas, durante un breve período, resplandecen con el mismo brillo que galaxias enteras. Propusieron que tales explosiones se producen por el colapso de un núcleo estelar, excesivamente masivo, en la última etapa de su evolución, y que el colapso causa que los electrones y los núcleos de hidrógeno (protones) se combinen dando neutrones, creando así una estrella de neutrones. Dado que los neutrones respetan a su vez el principio de exclusión de Pauli, pueden, también ellos, asumir un estado degenerado y, por tanto, pueden evitar que la estrella siga colapsando. Una estrella de neutrones de una masa solar se hace estable a un radio de unos 10 kilómetros y a una densidad de unos 10^{14} gramos por centímetro cúbico.

La energía potencial gravitatoria liberada por la contracción de materia, durante la formación de una estrella de neutrones, se transforma parcialmente en energía cinética de rotación, al girar cada vez más rápido la estrella que colapsa a fin de conservar su momento angular. Una cantidad parecida de energía se escapa en una oleada de neutrinos y en la expansión catastrófica de la envoltura externa de la estrella, la cual forma la típica nube en expansión de un residuo de supernova. La energía almacenada en la rotación de la estrella de neutrones escapa gradualmente en corrientes de partículas de alta energía.

De alguna manera que todavía no se comprende, las corrientes de partículas generan haces de radiación electromagnética que, al girar la estrella de neutrones, se observan como la señal periódica característica de un pulsar. Durante los últimos 10 años se han descubierto unos 150 pulsars. El más rápido gira 30 veces por segundo; se trata del residuo estelar de la supernova que hizo surgir la Nebulosa del Cangrejo. Girando a esa velocidad, cualquier objeto retenido sólo por

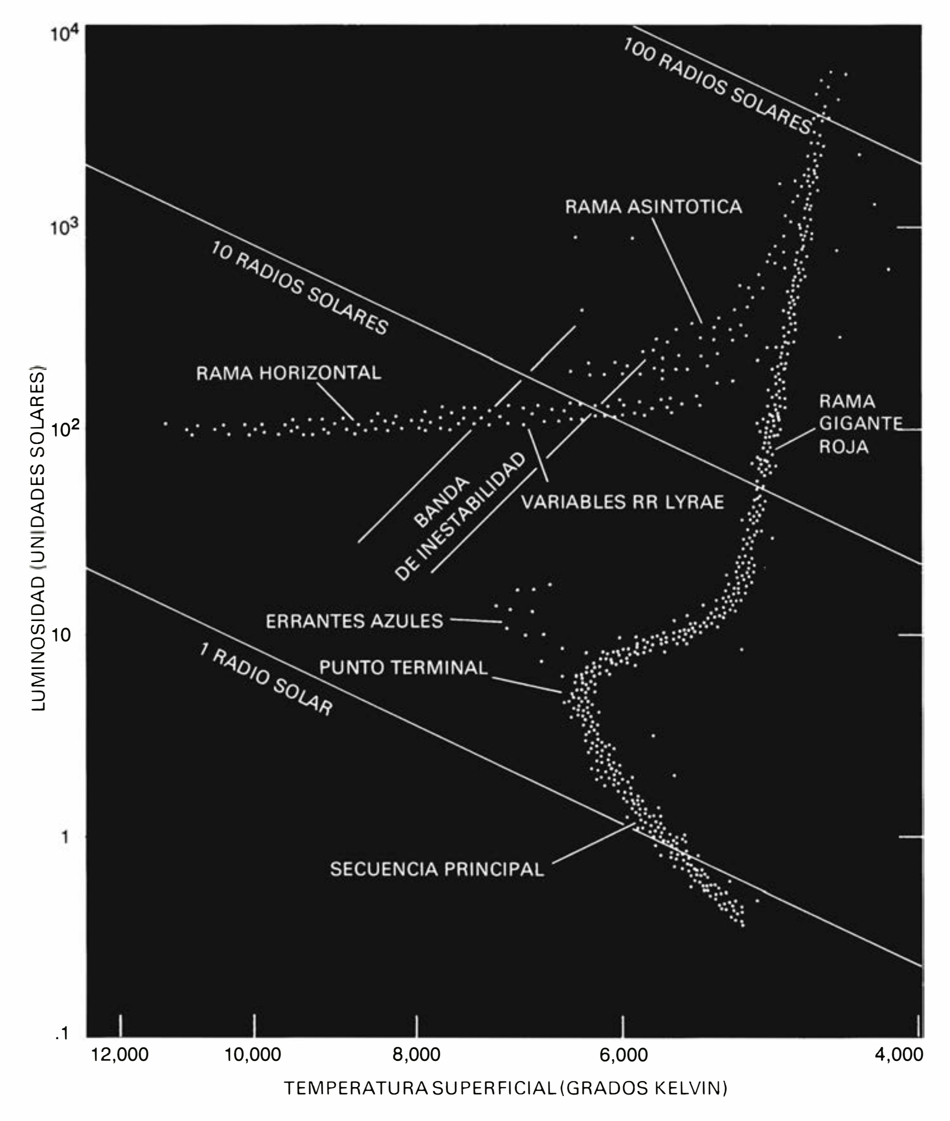
su propia gravedad se desintegraría, a menos que su densidad excediera los 10^{12} gramos por centímetro cúbico. (Los objetos terrestres, verbigracia, peonzas y ruedas voladoras, que dan vueltas a una velocidad mayor, están retenidos por fuerzas electromagnéticas, que son muy superiores). Puesto que 10^{12} gramos por centímetro cúbico es más de un millón de veces la densidad de una enana blanca típica, no cabe la menor duda de que el pulsar del Cangrejo es una estrella de neutrones. Si se considera el ritmo al que tienen lugar estos acontecimientos en nuestra galaxia, se llega a la conclusión de que varios cientos de millones de estrellas de neutrones se han formado y están ahora paseándose por el disco galáctico, invisibles a excepción de las pocas que se hallan a una proximidad suficiente y son lo bastante jóvenes para que puedan detectarse como pulsars.

El comportamiento de la materia en aquellos conglomerados de los cuales se ha extraído toda la energía nuclear disponible puede esquematizarse representando el radio de un cuerpo esférico en función de su masa *(véase la ilustración de la página 23)*. Para simplificar, se pueden considerar conglomerados de hierro, el cual cae al fondo de la curva de energía de ligadura nuclear; ello indica que no puede extraerse ya más energía nuclear sean cuales fueren la temperatura y la presión. Desde las balas "BB" de hierro hasta los planetas también de hierro, hay poca desviación con respecto a la curva de densidad constante. Pero cuando la masa excede 0,01 masas solares, se da una compresión sustancial, y el radio decrece efectivamente de unos 20.000 kilómetros a unos 10.000. Este es el intervalo de masas de las enanas blancas. En el límite de Chandrasekhar, que para el hierro es de 1,2 masas solares, la curva es horizontal y no hay radio estable por encima ni a la izquierda. Un núcleo estelar degenerado que crezca por encima de esa masa deberá colapsar hacia el régimen de estrellas de neutrones descrito por una curva que va hacia abajo, desde un radio de unos centenares de kilómetros hasta otro de menos de 10 kilómetros; lo cual se corresponde con un aumento de la masa desde 0,1 masas solares hasta varias veces la masa del Sol.

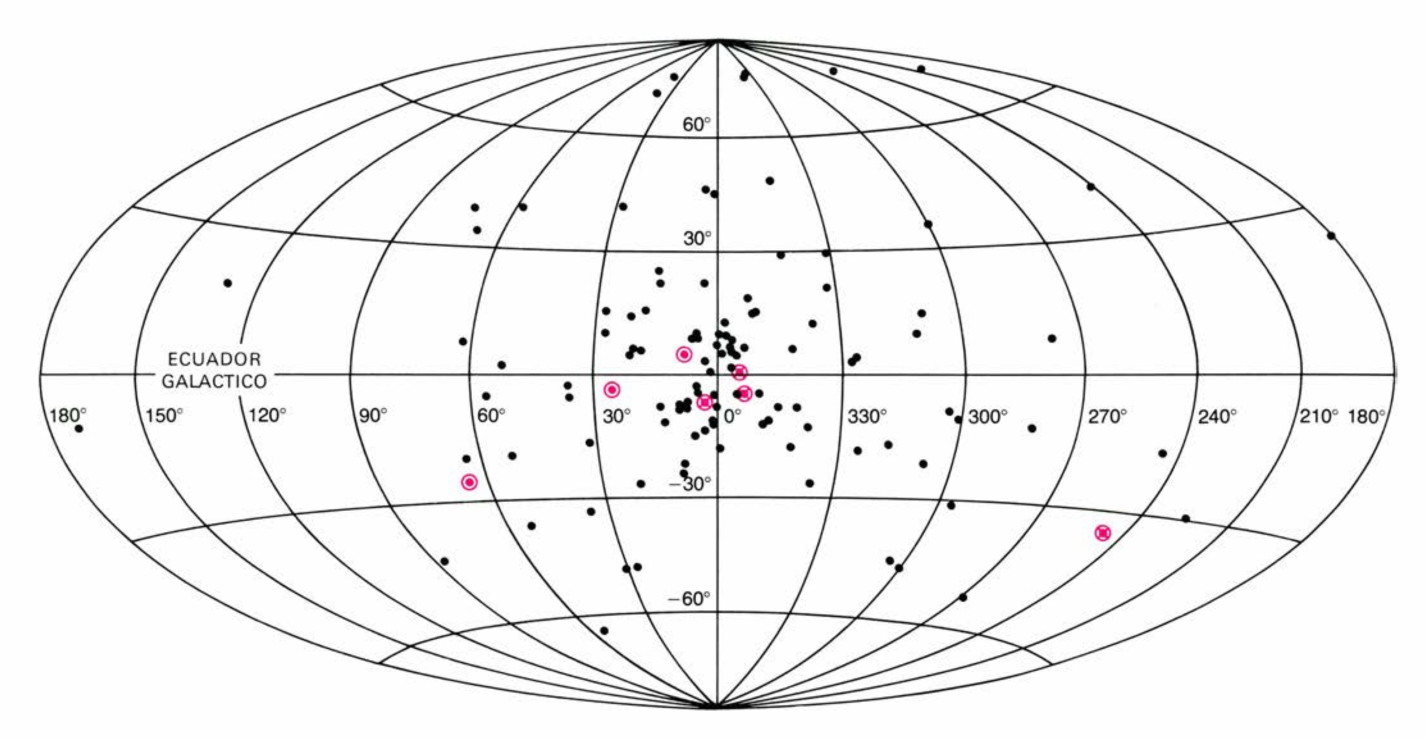
Por encima de dos masas solares, sólo hay extensiones, no confirmadas, de la teoría. Los cálculos de J. Robert Oppenheimer y G. M. Volkoff, basados en la teoría general de la relatividad, colocan un límite superior de 0,7 masas solares a la masa de una estrella degenerada que esté compuesta de neutrones sin interacción. Cálculos más recientes, que tienen en cuenta las interacciones de los neutrones, elevan el límite superior hasta unas tres masas solares. La teoría más convincente predice que una estrella degenerada, cuya masa exceda ese límite y que esté totalmente desprovista de energía nuclear, colapsará a un agujero negro, con un radio en kilómetros que es aproximadamente igual a tres veces su masa, expresada en masas solares.

Este es, pues, el esquema general de la evolución estelar que puede conducir a la formación de enanas blancas, estrellas de neutrones aisladas y, posiblemente, agujeros negros. Antes de pasar a considerar cómo se hallan implicados tales objetos en la formación de estrellas de rayos X, debemos examinar dos importantes cuestiones. ¿Qué edades tienen las estrellas de los cúmulos? ¿Cuál es el intervalo de masas para las estrellas de la secuencia principal que evolucionan hacia estrellas de neutrones?

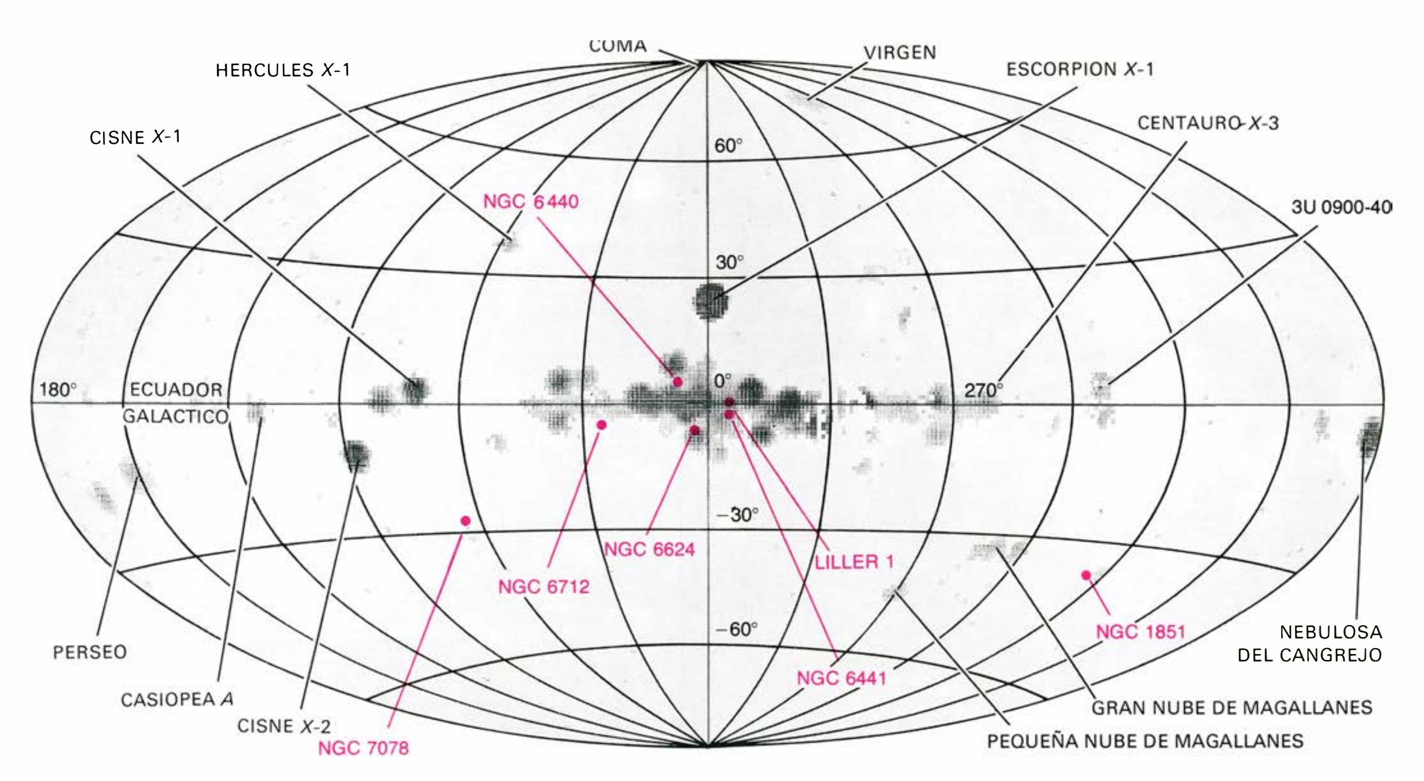
En la década de 1950, comenzaron a avanzarse las respuestas a la primera pregunta; los progresos de los ordenadores electrónicos hicieron posibles entonces los modelos numéricos de la estructura estelar. Puesto que no puede observarse directamente el interior de una estrella (excepto, tal vez, el del Sol, cuyos procesos internos se están estudiando a través de los ensayos de medición del flujo de neutrinos solares), deben extraerse de esos modelos predicciones de apariencias externas que puedan con-



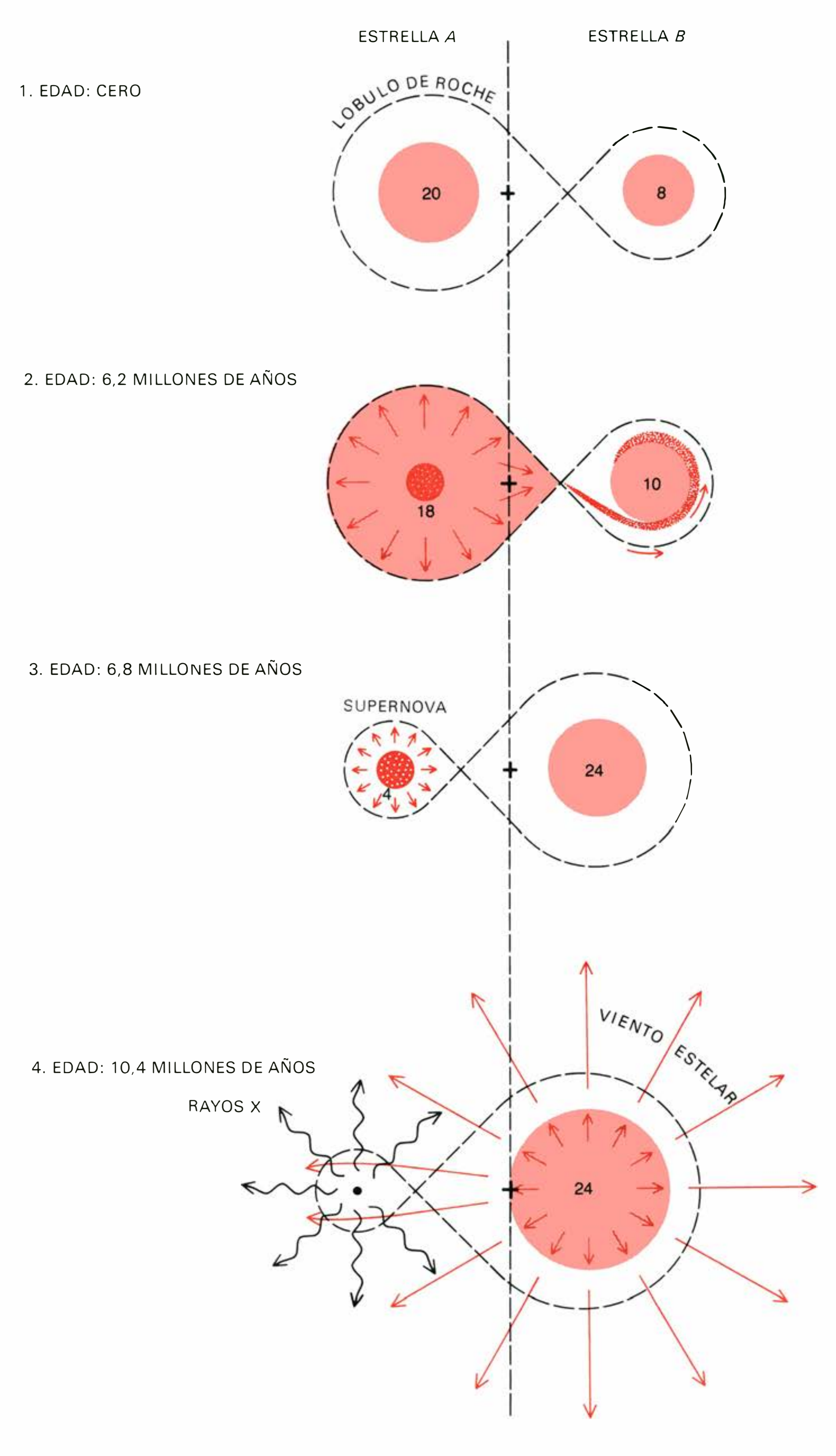
LA DISTRIBUCION TEMPERATURA-LUMINOSIDAD de las estrellas de un cúmulo globular típico viene representada en un diagrama de Hertzsprung-Russell (H-R). Una estrella comienza su vida en la "secuencia principal" del diagrama, "quemando" por reacciones termonucleares el hidrógeno de su núcleo. Permanece en la secuencia principal en tanto dura la provisión de hidrógeno de su núcleo. Cuando se acaba el hidrógeno, al haberse convertido en helio por las reacciones termonucleares, la estrella se aparta de la secuencia principal y atraviesa rápidamente, hacia arriba, la rama gigante roja del diagrama, quemando el hidrógeno que queda en su capa externa. En el extremo de la rama gigante roja, el helio del núcleo de la estrella está lo bastante caliente como para arder; genera un violento "centelleo del helio", después del cual la estrella aparece en la rama horizontal. Más tarde, entrará en la rama asintótica, quemando hidrógeno y helio en capas separadas. Muchas estrellas expulsan capas exteriores y terminan como enanas blancas o estrellas de neutrones. Algunas de ellas pueden convertirse en agujeros negros.



DISTRIBUCION DE LOS CUMULOS GLOBULARES, en un mapa completo del cielo, representado en coordenadas galácticas. Se conocen unos 130 cúmulos globulares. Es de suponer que cerca de 70 más estén ocultos por el polvo del disco galáctico. Los siete cúmulos que tienen estrellas de rayos X están rodeados por un círculo en color; con una cruz, los cuatro en los que se han detectado erupciones de rayos X.



MAPA DEL CIELO EN RAYOS X, hecho con ordenador a partir de los datos obtenidos con el detector de rayos X del Instituto de Tecnología de Massachusetts a bordo del Seventh Solar Orbiting Observatory (OSO-7). El detector tiene un campo visual de tres grados, con lo que las imágenes de las fuentes puntuales aparecen como círculos difusos con un radio de tres grados. Muchas de las fuentes de rayos X son sistemas binarios de la galaxia propiamente dicha, en los cuales un miembro es una estrella de neutrones o, posiblemente, un agujero negro, como por ejemplo Hércules X-1, Escorpión X-1, Cisne X-1, Cisne X-2, Centauro X-3 y 3U 0900-40. La Nebulosa del Cangrejo y Casiopea A son residuos nebulosos de supernovas que emiten rayos X; también están dentro de nuestra galaxia. Se han detectado asimismo fuentes de rayos X en dos galaxias cercanas: la Gran y la Pequeña Nube de Magallanes. En las constelaciones Virgen, Perseo y Cabellera de Berenice hay cúmulos de galaxias, alejadas 30 a 200 millones de años-luz, que contienen galaxias y materia intergaláctica emisoras de rayos X detectables. Las seis fuentes rotuladas con números NGC están en cúmulos globulares de nuestra galaxia, como ocurre también con Liller 1.



LA EVOLUCION DE UN SISTEMA BINARIO CERRADO sigue, según se cree, el esquema aquí descrito. Al principio, las estrellas *A* y *B*, ambas todavía en la secuencia principal, son miembros masivos de un sistema binario típico. Debido a que *A* posee una masa mucho mayor (20 masas solares) que la estrella *B* (ocho masas solares), consume rápidamente el hidrógeno en su núcleo, empieza a quemar helio, se aleja de la secuencia principal y se expande a gigante roja. Tras 6,2 millones de años, masa de la estrella *A* empieza a desbordar su lóbulo de Roche (el límite hasta el cual su materia está gravitatoriamente ligada) y es capturada por la estrella *B*. A los 6,8 millones de años, la estrella *A* estalla como una supernova, dejando una estrella de neutrones. Después de 10,4 millones de años, la estrella *B* se ha convertido en una supergigante y comienza a emitir masa en forma de un fuerte "viento" estelar, parte del cual queda capturado por la estrella de neutrones. La materia que cae cede energía en forma de rayos X. Este esquema está basado en cálculos realizados por Camiel de Loore y Jean Pierre de Grève, de la Universidad de Bruselas.

trastarse con las propiedades observables de las estrellas: masa, luminosidad, temperatura superficial, composición de la superficie y edad. Los lugares ideales para comprobar tales hipótesis de trabajo son los cúmulos estelares, tanto los viejos cúmulos globulares que se encuentran principalmente en el halo galáctico como los cúmulos más jóvenes, más pequeños y más bien amorfos que se encuentran en las regiones de activa formación estelar del disco galáctico.

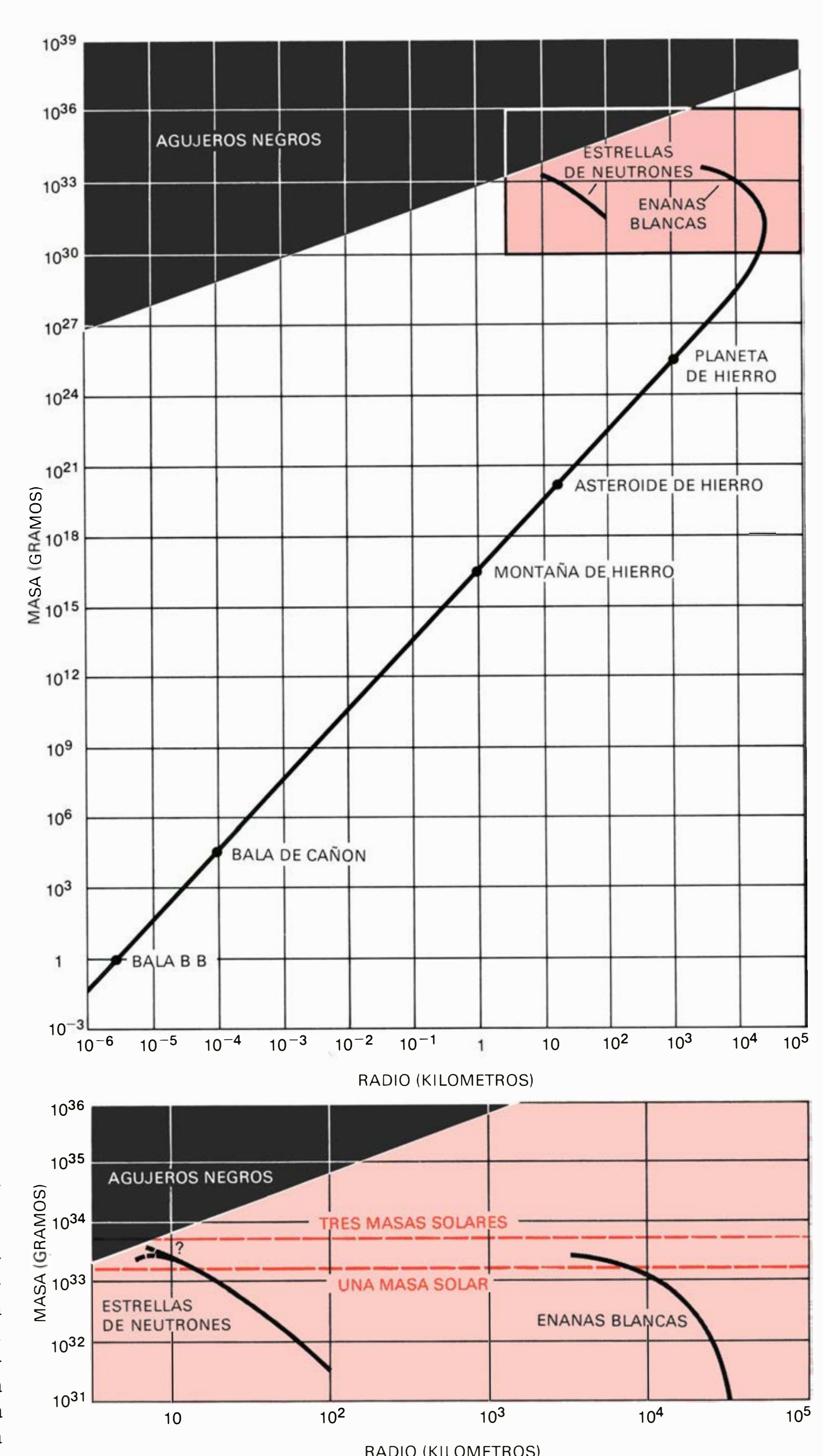
Para nuestro propósito, la propiedad a destacar en un cúmulo es que sus estrellas se formaron todas casi al mismo tiempo, a partir de un material cuya composición era aproximadamente la misma. Así pues, las diferencias entre las características externas observadas de las estrellas del cúmulo se deben, en su mayor parte, a diferencias en sus masas iniciales. Puesto que el ritmo de evolución aumenta con la masa, el diagrama H-R de un cúmulo es, en efecto, una instantánea de los estados evolutivos de multitud de estrellas de la misma edad y composición, aunque con distintas masas iniciales. Para someter a prueba un modelo, se verifica si el diagrama H-R de estrellas calculadas con varias masas iniciales puede adaptarse a los diagramas H-R de cúmulos reales, ajustando la edad y la composición inicial. Los cálculos llevados a cabo, en la década de 1960, por Icko Iben, Jr., y Robert T. Rood mostraron que el diagrama H-R de estrellas modelo se adecuaba satisfactoriamente al diagrama H-R de estrellas de cúmulos globulares típicos, suponiendo que las estrellas tenían una edad de 13.000 millones de años y una composición inicial de hidrógeno mezclado con un 29 por ciento de helio y un 0,1 por ciento de elementos más pesados. Según el modelo de Iben y Rood, las estrellas de cúmulos globulares de 0,8 masas solares alcanzan el punto de salida de la secuencia principal a una edad justo por debajo de los 10.000 millones de años.

Los problemas que restan en el ámbito teórico y en el cálculo dejan todavía muchos cabos por atar en lo que respecta al intervalo de masas iniciales de las estrellas progenitoras de las estrellas de neutrones. Se ha deducido un límite inferior a partir de la observación de existencia de enanas blancas en las Híades, un cúmulo galáctico cercano cuyos miembros más brillantes forman la *V* de la constelación Tauro. En este cúmulo se encuentran estrellas de la secuencia principal de hasta cuatro masas solares, lo cual implica que las enanas blancas descienden de estrellas cuyas masas iniciales sobrepasaban en algo ese valor de cuatro masas solares. Los cálcu-

los de modelos muestran que la vida de quemadora de combustible nuclear, de una estrella de cuatro masas solares, es inferior a los 100 millones de años.

Pasemos ahora a los sistemas binarios masivos "cerrados", en donde la transferencia de materia influye en la evolución de ambas compañeras. Al principio de sus vidas, las compañeras evolucionan como estrellas aisladas, con perturbaciones de menor cuantía, si es que se dan. Cada estrella está contenida en su propio pozo de potencial gravitatorio. Se produce, sin embargo, un cambio radical cuando la estrella *A*, el miembro de mayor masa en la pareja binaria cerrada, termina la etapa en la cual quema hidrógeno en el centro y empieza a quemarlo en su capa. Como hemos visto, la estrella, entonces, se sale de la secuencia principal y se expande mientras sube por la rama gigante roja del diagrama H-R. En algún momento, la envoltura externa de la estrella *A* desborda la barrera de potencial que la separa de su pareja, la estrella *B*, y se transfiere masa en gran escala a la estrella *B*. El resultado es que la estrella *A* se queda sin envoltura externa, manteniendo sólo su núcleo, altamente evolucionado, que continúa hacia los últimos estadios de evolución y explosión como supernova.

Mientras tanto, la estrella *B*, por heredar la envoltura de la estrella *A*, tiene ahora masa suficiente para mantener ligado el sistema a través de la eventual explosión de supernova. A partir de entonces, el sistema es una binaria que consta de la estrella *B* y el residuo de la estrella *A*, el cual es ahora una estrella de neutrones en la compañía de su, desde hace poco, masiva pareja. La estrella *B* sigue evolucionando; muy pronto comenzará a devolver materia a la estrella de neutrones, sea por emisión de partículas de alta velocidad (el "viento" estelar), sea expandiéndose hasta que su envoltura desborde su propia barrera gravitatoria. En este viaje de vuelta, la materia se verá arrastrada al pozo de potencial de la estrella *A*, el cual resultará entonces tan profundo que un protón en caída libre llegaría a la superficie de la estrella de neutrones con una energía cinética de más de 100 millones de electrón-voltios (100 MeV). Esta energía es 14 veces la energía liberada por protón cuando fusionan el hidrógeno y el helio. En efecto, el proceso es tan eficaz que un ritmo de acreción de sólo una masa solar cada mil millones de años basta para suministrar toda la potencia radiada por una típica estrella de rayos X de alta luminosidad. Además, a este bajo ritmo de captura, la densidad del material que



LA RELACION MASA-RADIO para una bola de hierro fría *(diagrama superior)* **sigue una línea recta hasta una masa de unos 10^{30} gramos. La región comprendida entre los 10^{31} y los 3×10^{33} gramos, que se muestra agrandada en el diagrama inferior, está ocupada por estrellas enanas blancas, las cuales están sostenidas frente al colapso por la presión de electrones en estado "degenerado", concepto mecánico-cuántico que explica por qué electrones densamente empaquetados, con cierta cantidad limitada de energía, resisten nuevas compresiones. La adición de masa por encima de 3×10^{33} gramos (1,4 masas solares) es causa de que los electrones se combinen con protones para formar neutrones, dando así lugar a una estrella de neutrones. Esta también se encuentra apoyada contra el colapso por la degeneración de sus partículas constituyentes. La teoría indica que no hay estado estable para objetos por encima de tres masas solares y que la adición de masa, por encima de un valor crítico, provocará que la estrella de neutrones colapse a agujero negro. El diagrama de la ilustración inferior está basado en un estudio de los estados densos de la materia por Steven Weinberg, que se halla actualmente en la Universidad de Harvard.**

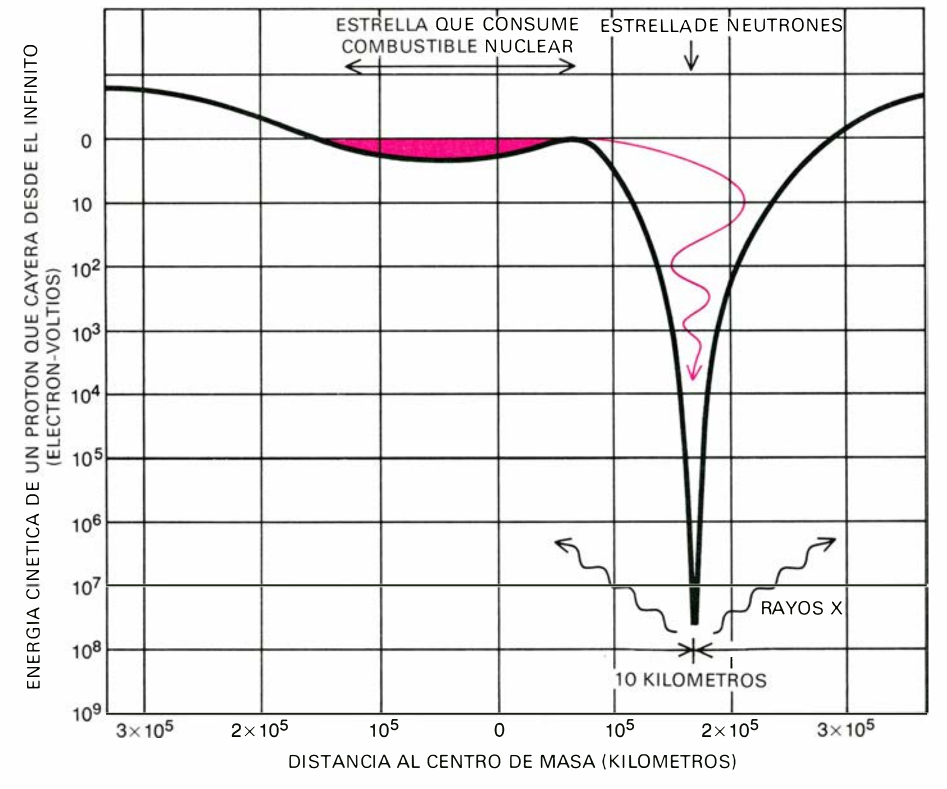
cae resulta tan exigua que permite que los rayos X emitidos cerca de la estrella de neutrones se escapen del sistema con pocas pérdidas por absorción durante su salida. Montajes como el expuesto fueron estudiados de modo intensivo por Edward P. J. van den Heuvel y sus colaboradores de la Universidad de Amsterdam, y por otros grupos. Su trabajo suministra explicaciones plausibles para una amplia variedad de estrellas de rayos X que se encuentran fuera de los cúmulos globulares.

En realidad, el flujo de la materia a capturar hacia una estrella de neutrones queda retardado por ondas de choque, campos magnéticos, fuerzas centrífugas y presión de radiación. En vez de caer radialmente hacia dentro, la materia forma un "disco de captura" en rotación de plasma caliente, que es un gas de partículas cargadas. La materia describe una espiral hacia dentro, a través del disco de captura, y finalmente se ve conducida a la superficie en los polos magnéticos de la estrella de neutrones; con ello cada protón pierde su energía, no en un solo rayo gamma de 100 MeV, sino por muchos rayos X, principalmente en el intervalo por debajo de los 20 KeV. Los polos magnéticos no suelen estar sobre el eje de rotación, por lo que las regiones de emisión intensa de rayos X giran fuera de eje. Resulta así que la emisión de rayos X,cuando se la observa a distancia, parece estar pulsando periódicamente. El estudio detallado de estas pulsaciones no deja ninguna duda respecto de que las fuentes de rayos X sean estrellas de neutrones magnetizadas y en rotación.

La presión de la radiación que fluye a partir de una estrella de neutrones que captura materia controla el propio ritmo de la captura. La fuerza de la gravedad sobre un plasma que fluya uniformemente hacia dentro queda equilibrada por la presión de radiación, si una estrella de una masa solar es 50.000 veces más luminosa que el Sol. A una luminosidad mayor, tal estrella expulsaría su capa más externa. La limitación que este efecto impone a la luminosidad de estrellas masivas fue apuntada por A. S. Eddington en la década de 1920. La relevancia del límite de Eddington para el ritmo de captura de las estrellas de rayos X se complica por la probable canalización de los flujos de captura y por las variaciones direccionales en la emisión de rayos X. Sin embargo, las estrellas de rayos X más luminosas tienden a apiñarse en el límite de Eddington para estrellas de una masa solar, lo cual indica que están capturando materia a un ritmo que se halla muy cerca del máximo permitido. Si el ritmo de captura excede el límite al cual puede escaparse la radiación generada o al que puede expulsarse el exceso de materia, ésta se amontonará en torno a la estrella de neutrones y extinguirá su emisión de rayos X.

La idea de que las estrellas de rayos X son estrellas de neutrones, que están capturando materia en sistemas binarios, fue avanzada por vez primera por I. S. Shklovskii, en 1967, en un intento por explicar los fenómenos asociados a Sco X-1; ésta es una estrella de rayos X en la constelación del Escorpión, descubierta cinco años antes en una observación desde cohete hecha por Riccardo Giacconi y sus asociados, de la American Science and Engineering, Inc. En 1965, el grupo de la American Science and Engineering, que trabajaba en colaboración con el grupo de astronomía de rayos X del Instituto de Tecnología de Massachusetts y astrónomos ópticos de los Hale Observatories y de Tokio, habían identificado la contrapartida óptica de Sco X-1 como una tenue estrella variable azul. Sin embargo, no se percibieron señales claras de un sistema binario, ni en el comportamiento en rayos X ni en el óptico, de Sco X-1, hasta muy recientemente. Mientras tanto, el First Small Astronomy Satellite *(Uhuru)*, lanzado en 1970 y que llevaba incorporado un observatorio de rayos X, desarrollado bajo la dirección de Giacconi, descubrió cuatro fuentes binarias de rayos X. Eran inconfundibles, puesto que una compañera eclipsaba regularmente a la otra cuando ambas se movían una entorno a la otra. Además, dos de las binarias pulsaban con períodos que mostraban variaciones Doppler debidas al movimiento orbital de la binaria. En la actualidad se conocen más de una docena de binarias pulsantes de rayos X, con una notable variedad de ritmos de pulsación, períodos orbitales, características del eclipse y variaciones de luminosidad. Tal diversidad se explica, al parecer, en función de los intervalos disponibles de las características iniciales para los sistemas binarios, a saber, las masas de las dos compañeras y sus elementos orbitales.

Una peculiaridad compartida por todas las binarias eclipsantes de rayos X es su juventud. Ninguna puede ser mucho mayor que la suma de las edades en la secuencia principal de dos estrellas; una de ellas tiene una masa inicial lo bastante grande (más de cuatro masas solares) para dar lugar a una estrella de neutrones y, la otra, una masa inicial eficaz que equivale a aquella que ha ob-



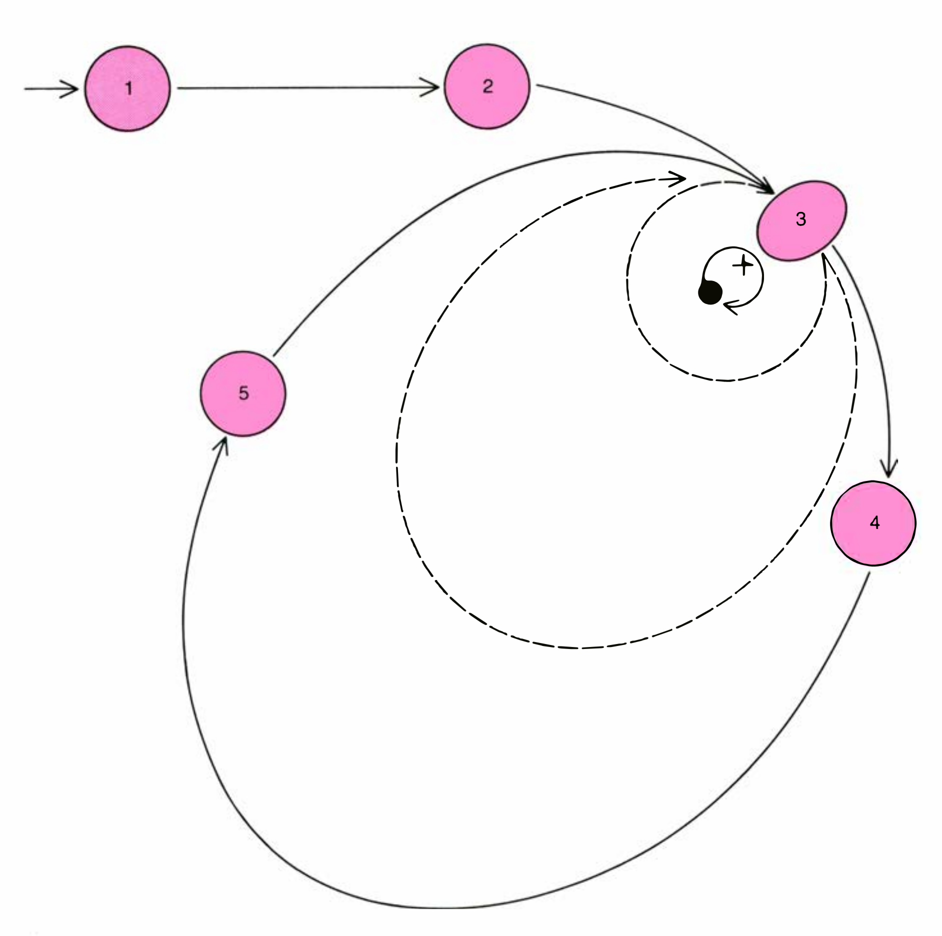
LA REPRESENTACION DE LA ENERGIA POTENCIAL de una binaria de rayos X describe dos pozos de potencial: uno ancho y poco profundo, que contiene la masa de la estrella que está todavía en la etapa de quema del combustible nuclear, y otro estrecho, hondo, que contiene la masa de la estrella de neutrones compañera. Toda materia que caiga, desde la superficie de la estrella grande en el pozo de potencial de la estrella de neutrones, adquiere 100 millones de electrón-voltios de energía cinética por protón, la cual puede liberarse en calor y radiarse en fotones de rayos X, con energías de 1000 a 100.000 electrón-voltios. Los protones caen en espiral.

tenido la secundaria al final del primer episodio de transferencia de masa. En general, esto implica una edad inferior a unos cientos de millones de años, por lo que tales sistemas no podrían haber sobrevivido los 13.000 millones de años que han transcurrido desde que empezaron a formarse las estrellas en los cúmulos globulares. Por tanto, la serie de acontecimientos en la evolución binaria que suministra una explicación tan plausible de la variedad y frecuencia de las estrellas de rayos X en el disco galáctico, no es válida para las estrellas de rayos X de los cúmulos globulares, salvo, tal vez, para un intervalo extremadamente reducido e improbable de condiciones de partida.

Me di cuenta de este dilema en 1975, cuando mi colega, entonces estudiante postgraduado, Thomas H. Market, descubrió que el mapa de exploración del cielo en rayos X que estaba preparando a partir de los datos de satélites, mostraba una fuente de rayos X, no detectada anteriormente, en una posición alejada del plano galáctico y que coincidía con el cúmulo globular NGC 1851. Los datos se habían obtenido con un detector de rayos X, desarrollado por el grupo de astronomía de rayos X del M.I.T., para el Seventh Orbiting Solar Observatory, que fue lanzado en septiembre de 1971.

Uno de nuestros objetivos era preparar un mapa completo del cielo a las longitudes de onda correspondientes a los rayos X. Una exploración similar se había llevado a cabo con el *Uhuru*, que salió publicada en 1974 en forma de una extensa lista de fuentes de rayos X: el catálogo "3U". En la lista había diversas fuentes variables identificadas con cúmulos globulares. Nuestra exploración con el OSO-7 confirmó tres de ellas y descubrió otras dos más.

Esos cinco objetos constituían claramente una clase de estrellas de rayos X que no encajaba en el esquema evolutivo que daba razón de las otras binarias de rayos X. Jonathan Katz, del Institute for Advanced Study, había señalado anteriormente que la proporción de luminosidad de rayos X con respecto a la masa conjunta de todos los cúmulos globulares era, por lo menos, dos órdenes de magnitud mayor que la misma proporción calculada para toda la galaxia. Ahora podríamos contar cinco fuentes de rayos X entre las 10^7 estrellas que se estiman para los cúmulos globulares, comparadas con las aproximadamente 100 fuentes de rayos X entre el total de 10^{11} estrellas de la galaxia entera. Así pues, la aparición preferente de estrellas de rayos X en la población de los cúmulos globulares, que ya hice notar al principio del artículo, se hizo aún más impresionante. El problema estribaba, por tanto, en hallar no sólo una nueva explicación para el origen de las estrellas de rayos X en los cúmulos globulares, sino en que además ella diera cuenta de tal frecuencia relativa de aparición, mucho mayor.

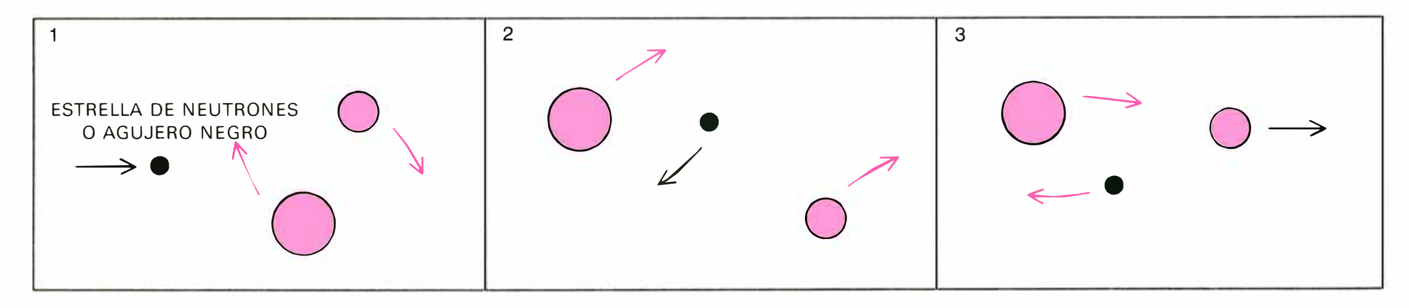


MODALIDAD DE FORMACION DE ESTRELLAS BINARIAS en los núcleos densos de cúmulos globulares, que visualiza la captura de una típica estrella quemadora de combustible nuclear *(color)* **por una estrella de neutrones o un agujero negro. Durante un encuentro cerrado de las dos estrellas, las mareas suscitadas disipan la suficiente energía mecánica de la pareja como para dejar ambos objetos ligados en órbita. Pérdidas por marea en sucesivos acercamientos hacen que la órbita, inicialmente excéntrica, se haga más y más circular. Andrew Fabian, James Pringle y Martin Rees, de la Universidad de Cambridge, propusieron este tipo de captura.**

La noción de estrella de rayos X como residuo estelar ultradenso, alimentado por captura, se impuso de tal modo que era razonable suponer que el mismo mecanismo básico estaba involucrado en las fuentes de rayos X de los cúmulos globulares. Las estrellas de masa mayor y de vida corta que nacieron cuando se formaron los cúmulos deben haber dejado muchas estrellas de neutrones y, posiblemente, agujeros negros. Cabe que algunas estrellas de neutrones salieran expulsadas del cúmulo por la violencia de la supernova que las creó, pero, sin lugar a dudas, muchas quedaron gravitatoriamente ligadas, aunque tal vez, inicialmente, en órbitas de gran amplitud.

En general, la energía mecánica de las estrellas de un cúmulo globular tiende a distribuirse por igual entre todas las estrellas, como lo hace entre las moléculas de un gas en un recipiente cerrado. Además, cuanto mayor sea la masa de una estrella con una energía dada, tanto menor será la distancia media entre la estrella y el centro del cúmulo. Por tanto, los residuos, estrellas de neutrones y agujeros negros, por estar dotados de masas mayores que la media de las estrellas, deben haberse congregado cerca del centro del cúmulo, una vez que sus energías hubieran alcanzado el equilibrio. Allí han permanecido durante miles de millones de años, cual objetos obscuros en espera de transformarse en estrellas de rayos X al adquirir una fuente de materia para la captura.

La hipótesis más probable, en cuanto a la fuente del material a capturar es, en cierto sentido, la más obvia: el material deben aportarlo estrellas compañeras binarias que estén aún en la etapa de quema del combustible nuclear. Dado que, como hemos visto, tales compañeras no pueden haber sobrevivido desde la época en que nacieron las estrellas del cúmulo, hubieron de adquirirse por cap-



MECANISMO de captura alternativo, propuesto por Jack Hills, de Cambridge. Se parece a las colisiones de intercambio observadas en reacciones nucleares. **La estrella de neutrones o el agujero negro incidentes desplazan a una estrella de poca masa en un sistema binario. La estrella desalojada escapa.**

tura. Seis cúmulos, de los siete que contienen estrellas o eruptores de rayos X, o ambas cosas, se cuentan entre los 20 cúmulos con densidades centrales mayores; ello podría parecer que favorece los procesos de captura capaces de crear binarias en su núcleo.

En principio, dos estrellas pueden capturarse mutuamente a través de un encuentro muy cercano entre tres estrellas, en el cual una de ellas se lleve tanta energía cinética que las otras queden estrechamente ligadas. Pero, incluso en el núcleo más denso de un cúmulo globular, la frecuencia de los encuentros inmediatos entre dos estrellas es baja; no digamos nada de la frecuencia de los encuentros entre tres estrellas, de los que no cabe esperar que se dé ninguno en el curso de los 13.000 millones de años de vida de un cúmulo.

Para la creación de binarias por captura, se propuso muy pronto un par de mecanismos que sólo necesitan encuentros entre dos estrellas. Uno, avanzado por Andrew Fabian, James Pringle y Martin Rees, de la Universidad de Cambridge, se basaba en el hecho de que las estrellas de un cúmulo globular, consideradas como partículas gravitatoriamente atrapadas, se muestran relativamente "frías" en sus movimientos cinéticos. Por tanto, la disipación de una porción (aunque sea pequeña) de la energía cinética total de dos estrellas que experimentan un encuentro cercano puede resultar en su mutua captura. Si una estrella que esté quemando combustible nuclear tiene un encuentro cercano con una estrella de neutrones o con un agujero negro, se suscitará una marea en la estrella que esté quemando combustible nuclear, la cual disipará energía mecánica. Dentro de cierto intervalo de circunstancias, se disipará la energía suficiente para que ambos objetos queden gravitatoriamente ligados el uno al otro; al principio lo estarán en una órbita muy elíptica, con una separación pequeña en el punto de máximo acercamiento. Las pérdidas de energía mecánica por marea, en los sucesivos acercamientos cerrados, irán flexionando gradualmente la órbita hasta convertirla en un círculo casi perfecto, con un radio bastante corto para que ocurra una transferencia eficaz de masa en una época posterior. El escenario quedará a punto para una larga espera hasta que la estrella que quema combustible nuclear evolucione alejándose de la secuencia principal, se expansione y empiece a verter materia en el pozo gravitacional de su compañera, que antes era compacta y obscura.

El otro mecanismo de captura, elaborado por Jack Hills, de la Universidad de Cambridge, era análogo a las colisiones de intercambio que se explican en física atómica. Hills analizó, mediante simulaciones con ordenador, multitud de encuentros cercanos; por un lado, estudió un sistema binario cercano que constaba de dos estrellas de masa pequeña, típicas de la población de la secuencia principal en los cúmulos globulares, y, por otro, un residuo estelar pesado, como una estrella de neutrones o un agujero negro. Mostró que, a menudo, el resultado sería una colisión de intercambio, en la cual el residuo pesado sustituiría a una estrella de masa pequeña, se expulsaría la estrella sustituida fuera del sistema y quedarían en órbita la estrella de neutrones o el agujero negro con un aporte de materia para la captura.

Haciendo hipótesis plausibles acerca del número de estrellas de neutrones, tanto la teoría de la disipación por marea como la de las colisiones de intercambio predicen la formación de varias binarias cerradas, por captura, en los cúmulos globulares más condensados, donde la densidad estelar central excede las 1000 estrellas por año-luz cúbico. Una de esas binarias, al encenderse, podría emitir rayos X por un tiempo superior al de las binarias del disco galáctico, ya que su suministro de material para la captura proviene de una estrella de masa pequeña, que evoluciona lentamente y cuya fase de expansión, tras dejar la secuencia principal, dura cientos de millones de años. La lenta expansión permite que la materia se transfiera al modesto ritmo requerido para alimentar la emisión de rayos X sin ahogarla. Sobre esta base, las variaciones a largo término, observadas en estrellas de rayos X de los cúmulos globulares, se explican como resultado de una variabilidad del flujo de captura, debida a la excentricidad orbital residual o a una expansión no uniforme, del tipo de la que ocurre al final de la vida de las gigantes rojas. La aparente ausencia de pulsaciones, de las que cabría esperar que resultaran de la rotación de estrellas de neutrones, puede explicarse por la extinción de los campos magnéticos en estrellas de neutrones muy viejas. De hecho, nada impide pensar que la aparición de erupciones puede hallarse relacionada con la ausencia de campos magnéticos fuertes en estrellas tan antiguas.

Totalmente diferente fue la hipótesis ideada por Jeremiah P. Ostriker y John N. Bahcall, de la Universidad de Princeton, y por Joseph Silk y Jonathan Arons, de la Universidad de California en Berkeley. Propusieron que las fuentes de rayos X eran agujeros negros masivos, formados probablemente por la coalescencia de estrellas en el denso núcleo de un cúmulo que haya sufrido un colapso central catastrófico. Debe destacarse que las dificultades matemáticas y de cálculo para resolver un tal problema de cuerpos, resultan tan formidables que, ni siquiera en la actualidad, la pregunta de si puede tener lugar o no tal coalescencia catastrófica sigue sin hallar una respuesta tajante. Algunos investigadores creen que la primera binaria cercana masiva que se forme en el núcleo de un cúmulo actuará como una batidora y barrerá las estrellas circundantes, deteniendo así toda coalescencia antes de que ésta pueda progresar. En efecto, los miembros de un cúmulo están muy ligeramente ligados, hasta el punto de que un acortamiento de la separación orbital de una binaria cercana puede suministrar la energía mecánica suficiente para barrer todo un núcleo entero.

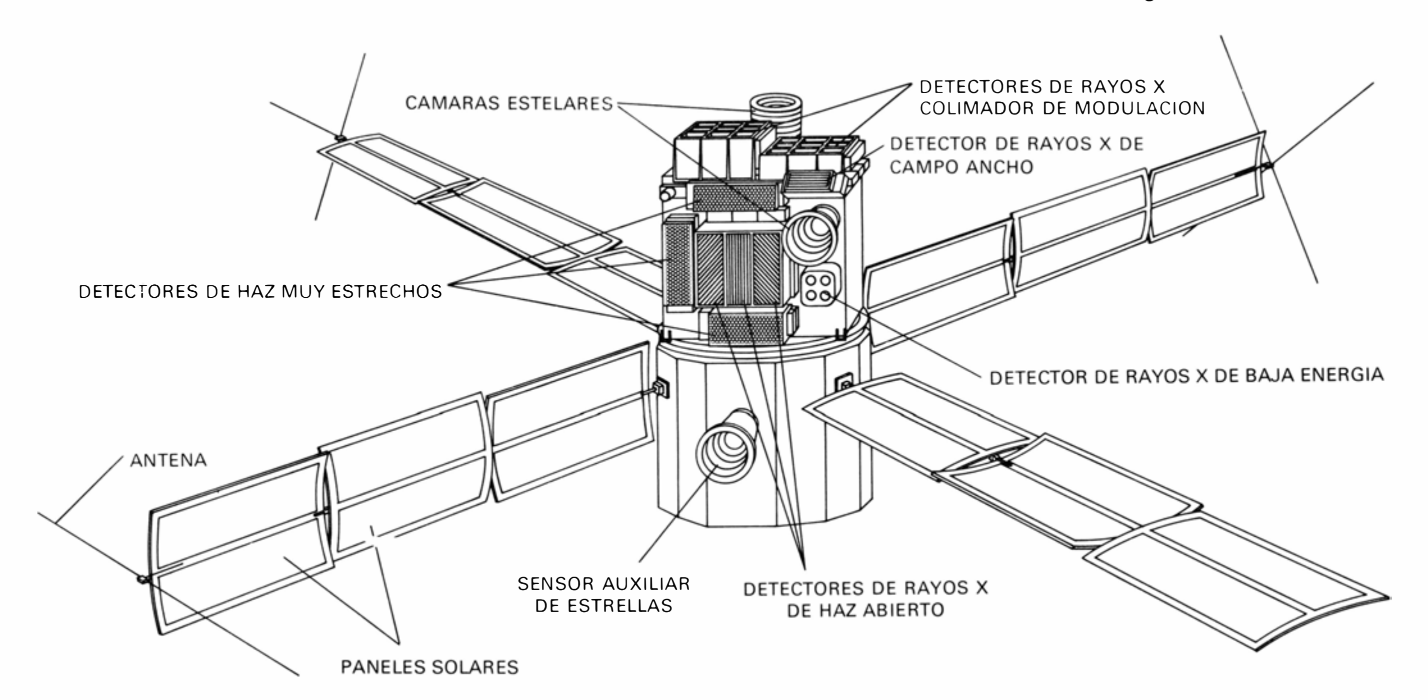
Los defensores de la hipótesis del agu-

jero negro masivo postulaban, sin embargo, la existencia de tales objetos en el centro de cúmulos altamente condensados. Muestran que el pozo de potencial de un tal cúmulo es bastante profundo como para retener el gas que fluye en vientos estelares y la materia que se pierde, en episodios esporádicos, por parte de estrellas normales que están quemando combustible nuclear. El material eyectado cae hacia el centro del cúmulo y alimenta el agujero negro a través del proceso de captura, generando con ello rayos X. Puesto que el material que cae es gas interestelar tenue, el agujero negro ha de ser muy masivo (comprendido entre 100 y 1000 masas solares) para poder arrastrarlo al ritmo que se necesita para generar la potencia en rayos X observada.

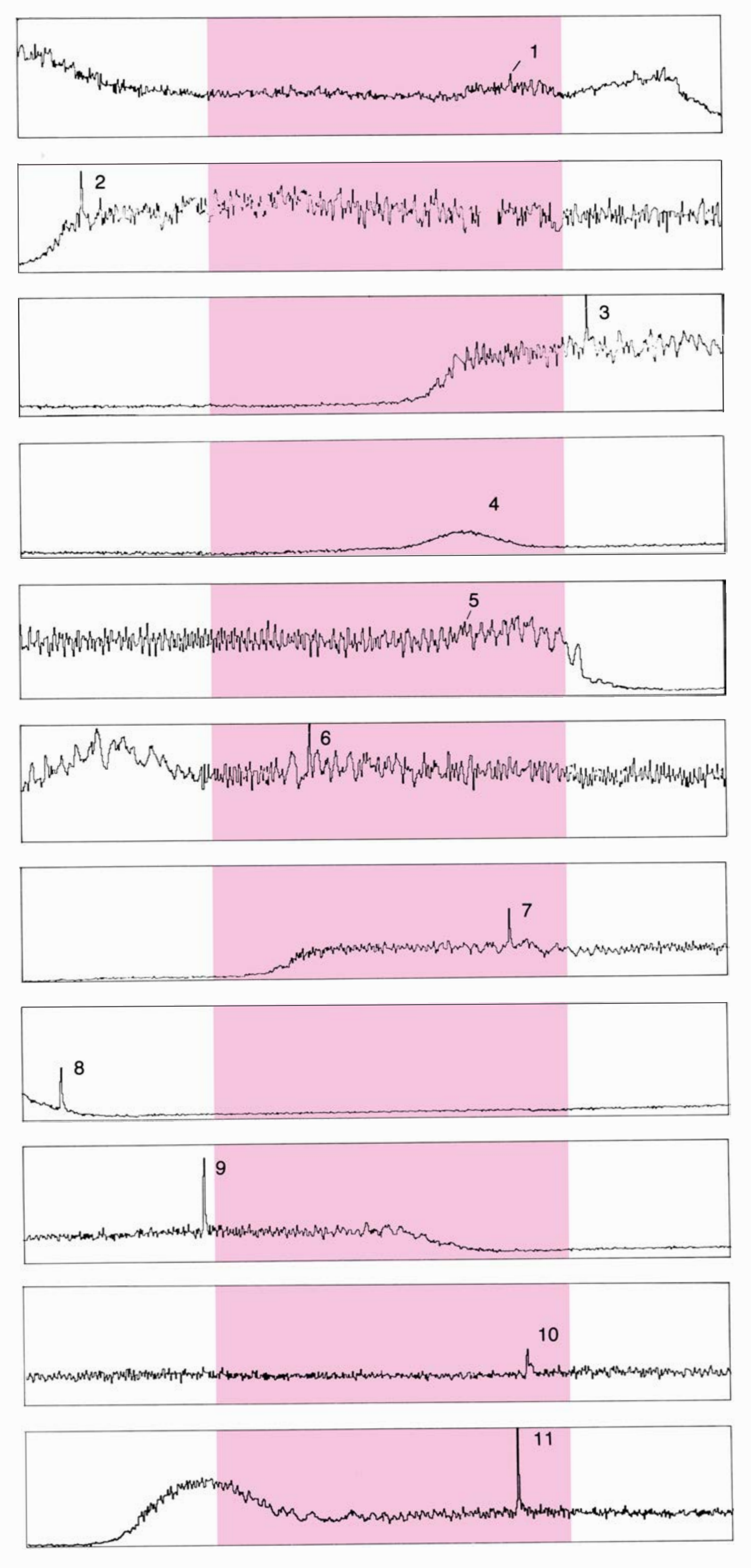
Así estaban las cosas cuando los astrónomos comenzaron a buscar pruebas que apoyaran una u otra de las dos hipótesis: el modelo de la binaria por captura y el modelo del agujero negro masivo, el "barredor del vacío". Pulsaciones periódicas o eclipses en una fuente de rayos X de un cúmulo globular apuntarían hacia el modelo de una binaria por captura, como fuente de rayos X. Nada de eso se ha observado. Un pico abrupto en la luminosidad general del cúmulo, situado junto al centro, favorecería la hipótesis del agujero negro; lo mismo ocurriría con la prueba de gas luminoso cerca del centro. Los datos, difíciles de adquirir, no son todavía decisivos. Sin embargo, la prueba de que las fuentes de rayos X en cúmulos ocurren preferentemente en los de densidades centrales elevadas es coherente con ambas hipótesis. En uno de los cúmulos globulares, NGC 6624, se mostró nítidamente la existencia de una densidad alta a partir de placas de corta exposición tomadas con el telescopio Wise de un metro, en Israel, por Neta A. Bahcall en 1975. En el núcleo del cúmulo hay un lóbulo impresionante de imágenes estelares irresolubles, con un diámetro de unos cuatro segundos de arco.

Inesperadamente, se registró un avance muy sugestivo. A finales de 1975, Jonathan E. Grindlay y Herbert Gursky, del Center for Astrophysics of the Smithsonian Astrophysical Observatory y del Harvard College Observatory, repasando datos del experimento de rayos X americano en el satélite astronómico holandés ANS, descubrieron la señal de una breve erupción de rayos X a partir del NGC 6624. Este cúmulo globular contiene la estrella variable de rayos X 3U 1820-30, que se había encontrado anteriormente en la exploración del *Uhuru;* se trata, además, de una de las fuentes variables más brillantes de entre las que están cerca del centro del disco galáctico. La posición definida por el *Uhuru* ofrecía la certeza de que 3 U 1820-30 se hallaba en el interior del cúmulo globular NGC 6624 y no era un objeto del disco. Claude R. Canizares, del grupo de astronomía de rayos X del M.I.T., trabajando con los datos de 1974 del OSO-7, encontró una ráfaga importante, de 20 minutos, en la intensidad de 3U 1820-30. Con el interés centrado ya en las estrellas de rayos X de los cúmulos globulares, era natural que 3U 1820-30 constituyera un blanco de primer orden en cuanto se dispuso de satélites capaces de apuntar a objetos específicos durante períodos largos. La estrella quedó incluida en el programa de observaciones del ANS, para marzo y septiembre de 1975. Fue uno de los primeros objetivos, para una medición refinada de la posición, que se marcó el observatorio de rayos X del M.I.T. a bordo del Third Small Astronomy Satellite (SAS-3), utilizando técnicas especiales desarrolladas bajo la dirección de Herbert W. Schnopper y de Hale V. Bradt. El SAS-3 se lanzó en mayo de 1975.

Las observaciones del ANS en el mes de marzo de 1975 encontraron a 3U 1820-30 en un estado de alta luminosidad media, sin variaciones espectaculares. Las observaciones del SAS-3 en el mes de mayo dieron en seguida una medición afinada de la posición, que redujo la incertidumbre a un círculo de error con un radio de 40 segundos de arco; el círcu-



EL TERCER SATELITE ASTRONOMICO (SAS-3) lleva un paquete de detectores de rayos X y de cámaras estelares diseñado y construido por el grupo de astronomía de rayos X del Instituto de Tecnología de Massachusetts. Las imágenes de los campos estelares suministradas por las cámaras permiten determinar la orientación del observatorio con una precisión de 30 segundos de arco en más o en menos. Por análisis con ordenador de los rayos X, registrados por detectores combinadores de modulación mientras el vehículo gira en torno al eje *Z*, pueden localizarse las fuentes de rayos X con una precisión de 20 segundos de arco, en más o en menos. El satélite se lanzó en el mes de mayo de 1975, desde una plataforma en el Ecuador, frente a las costas de Kenya. Maniobrada por el Centro de Investigación Aeroespacial de Italia, la plataforma había sido utilizada en 1970 para lanzar el pionero de los satélites astronómicos de rayos X, el *Uhuru*. El SAS-3 se gobierna por controladores del Goddard Space Flight Center, de la NASA, bajo la dirección, hora a hora, del grupo de astronomía de rayos X del M.I.T. Los datos, grabados en cinta, se envían desde el satélite a la tierra y se mandan órdenes desde tierra al satélite, cada 104 minutos.



lo comprendía el centro del cúmulo, en donde se esperaba que estuviera la estrella de rayos X. Durante la observación del SAS-3 en mayo, la luminosidad de 3U 1820-30 era baja, si la comparamos con el valor culminante encontrado por el *Uhuru*.

Por esas fechas, Garrett Jernigan, estudiante postgraduado que tenía bajo su responsabilidad desarrollar la mayor parte de las técnicas analíticas del proyecto SAS-3, estaba revisando los datos de "primera vista" que se recibían en el M.I.T.; tales datos se representaban en mapas configurados por ordenador a las pocas horas de que los registraran los detectores en órbita. Vio dos aumentos repentinos del ritmo de cuenta de rayos X, cada uno de los cuales duraba unos 10 segundos. El instrumento específicamente utilizado en la observación era un detector colimador de modulación, montado encima del satélite y con el eje orientado en la dirección de la rotación del satélite. El detector era sensible a rayos X con energías entre dos y 11 KeV. El campo visual definido por el colimador consistía en bandas alternadas de alta y baja sensibilidad, sobre un área rectangular de unos 24 grados de lado.

Al girar el satélite, las bandas de transmisión del colimador pasaban sobre las fuentes que hubieran en el ancho campo visual, modulando sus ritmos de cuenta en estructuras que podían reducirse, por análisis mediante ordenador, a gráficas de correlación que representaban cada fuente como un pico central con anillos concéntricos a su alrededor. Sin embargo, con tales erupciones brevísimas no pudo determinarse la posición de la fuente. Las erupciones hubieran podido proceder de cualquier punto del campo visual, de 24 por 24 grados, o incluso de

SERIE DE ERUPCIONES de rayos X registrada por el SAS-3 en mayo de 1975, procedentes de 3U 1820-30, la estrella de rayos X del cúmulo globular NGC 6624. Se registraron diez erupciones, con un intervalo medio de 4,4 horas. Es de suponer que fueran 11 erupciones en total; la cuarta de la serie tuvo lugar, al parecer, mientras la tierra bloqueaba la visual del satélite. Cada uno de esos 11 trazos contiene una porción de 40 minutos del registro de rayos X basado en 11 órbitas del SAS-3, centradas en el intervalo de recurrencia "esperado" de 4,4 horas. (La cuarta erupción esperada falta, evidentemente.) La porción coloreada del registro se extiende hasta 10 $^{1}/_{3}$ minutos a los lados del intervalo medio, y representa una desviación estándar (el "temblor") en la frecuencia media de las 10 erupciones registradas. Obsérvese que la erupción número 8 fue captada justo cuando la fuente estaba desapareciendo de la visual, detrás de la tierra. Se cambió la orientación del satélite, después de la erupción número 6, con el propósito de reducir la interferencia a partir de otras fuentes emisoras de rayos X.

algún "sobresalto" instrumental del satélite recientemente lanzado, en cuyo irreprochable funcionamiento estábamos empezando justamente a confiar. En todo caso, las diapositivas que preparó Jernigan para mostrar los datos que indicaban los dos aumentos bruscos se arrinconaron mientras nos afanábamos en otros descubrimientos, más fácilmente reconocibles, que nos iban llegando de las observaciones del SAS-3.

La observación del ANS se repitió en septiembre. Al inspeccionar los datos, unos meses más tarde, Grindlay y Gursky hicieron el sorprendente descubrimiento de la erupción de rayos X en NGC 6624. Tras una llamada telefónica de Grindlay, J. Heise, del laboratorio de investigación espacial perteneciente al Instituto de Astronomía de Utrech, encontró en seguida una segunda erupción. Ambas procedían de una estrecha franja del cielo que contenía 3U 1820-30 y el cúmulo globular. Los detectores eran sensibles a rayos X de uno a 30 KeV. Suponiendo que los rayos X se emitieran en todas direcciones por una fuente instalada en el cúmulo globular, la energía total de cada erupción era de más de 10^{39} ergios. Observando que las erupciones subían hasta la intensidad máxima en menos de un segundo, para luego decaer gradualmente durante los siguientes 10 segundos, Grindlay y Gursky sugirieron que los fotones de rayos X de una erupción se liberan en una "punta" de dos segundos y reverberan después al difundirse por una nube circundante de plasma caliente. La extinción gradual se interpretaba como resultado de la reverberación.

El modelo de Grindlay y Gursky suministraba también una explicación del hecho de que el espectro de las erupciones se hiciera más "duro", es decir, más rico en fotones de rayos X de mayor energía, al irse extinguiendo aquéllas. Postulaba que la nube circundante estaba tan caliente que, por término medio, los fotones difundidos ganaban energía. De acuerdo con la estimación de Grindlay y Gursky, la temperatura de la nube tenía que ser tan alta que sólo podría retenerla un cuerpo central de 1000 masas solares, el cual debería ser un agujero negro masivo. Trabajando independientemente en esa misma línea, Canizares mostró que una nube mucho más fría podría producir también la reverberación y el "endurecimiento" espectral observado, y requeriría, para mantenerse cohesionada, un cuerpo central de sólo una masa solar. En todo caso, el descubrimiento de este fenómeno, peculiar y totalmente inesperado, en un cúmulo globular intensificó grandemente el entusiasmo de muchos por la idea del agujero negro masivo.

El anuncio del descubrimiento de la erupción nos llevó a reexaminar nuestros datos procedentes del SAS-3. Pronto descubrimos que los dos aumentos repentinos señalados en el conjunto de datos de "primera vista" de Jernigan en mayo eran, en realidad, dos de una serie de diez erupciones registradas entre los datos completos "de producción" que se habían recibido en la National Aeronautics and Space Administration, varios meses después de la observación. Descubrimos con asombro que las erupciones eran recurrentes y casi periódicas, a un ritmo de una cada 4,4 horas; al parecer, faltaba una erupción en la serie cuando NGC 6624 era eclipsada por la tierra, al moverse el satélite en su órbita. Desconcertante y provocativo resultaba el hecho de que los intervalos entre erupciones sucesivas no fueran exactamente iguales; antes bien, poseían una dispersión de cerca del 5 por ciento. El fenómeno de la erupción tiene cuatro tiempos característicos, notablemente diferentes: un tiempo de subida de alrededor de un segundo, un tiempo de declive de unos 10 segundos, un tiempo de recuperación de aproximadamente 10^4 segundos y un "temblor", o variación en el tiempo de recuperación, de 600 segundos en más o en menos.

El temblor parecía especialmente significativo. Pensábamos si no sería producido por algo equivalente a un tornillo flojo en un mecanismo de relojería por lo demás exacto, o si no sería el resultado de alguna variabilidad en el nivel de puesta en marcha o en la fuente de potencia de un oscilador de amortiguamiento. John Bahcall y Ostriker, quienes atribuyeron las erupciones a los efectos de una estrella de neutrones en órbita en torno a un agujero negro masivo con disco de captura, propusieron inmediatamente un mecanismo del primer tipo. Sugirieron que la estrella de neutrones choca periódicamente con el disco de captura, generando pulsos, de período medio muy estable, si bien con un temblor producido por oscilaciones aleatorias del disco.

Tampoco se pensó dos veces el mecanismo del segundo tipo. Lo avanzó Fred Lamb, de la Universidad de Illinois; sugirió que un plasma que fluyera hacia la superficie de una estrella de neutrones magnetizada podría quedar atrapado en un "depósito" de la magnetosfera, donde se amontonaría hasta que rebasara cierta densidad crítica. Entonces, el depósito se derramaría y dejaría caer de repente un glóbulo de plasma sobre la superficie de la estrella, con lo que liberaría su energía gravitatoria en forma de una erupción de rayos X. El "escape" se cerraría luego y el "depósito" volvería a llenarse. En este modelo, el ritmo de recurrencia viene gobernado por diversidad de factores, entre ellos, el ritmo al que vuelve a llenarse el depósito.

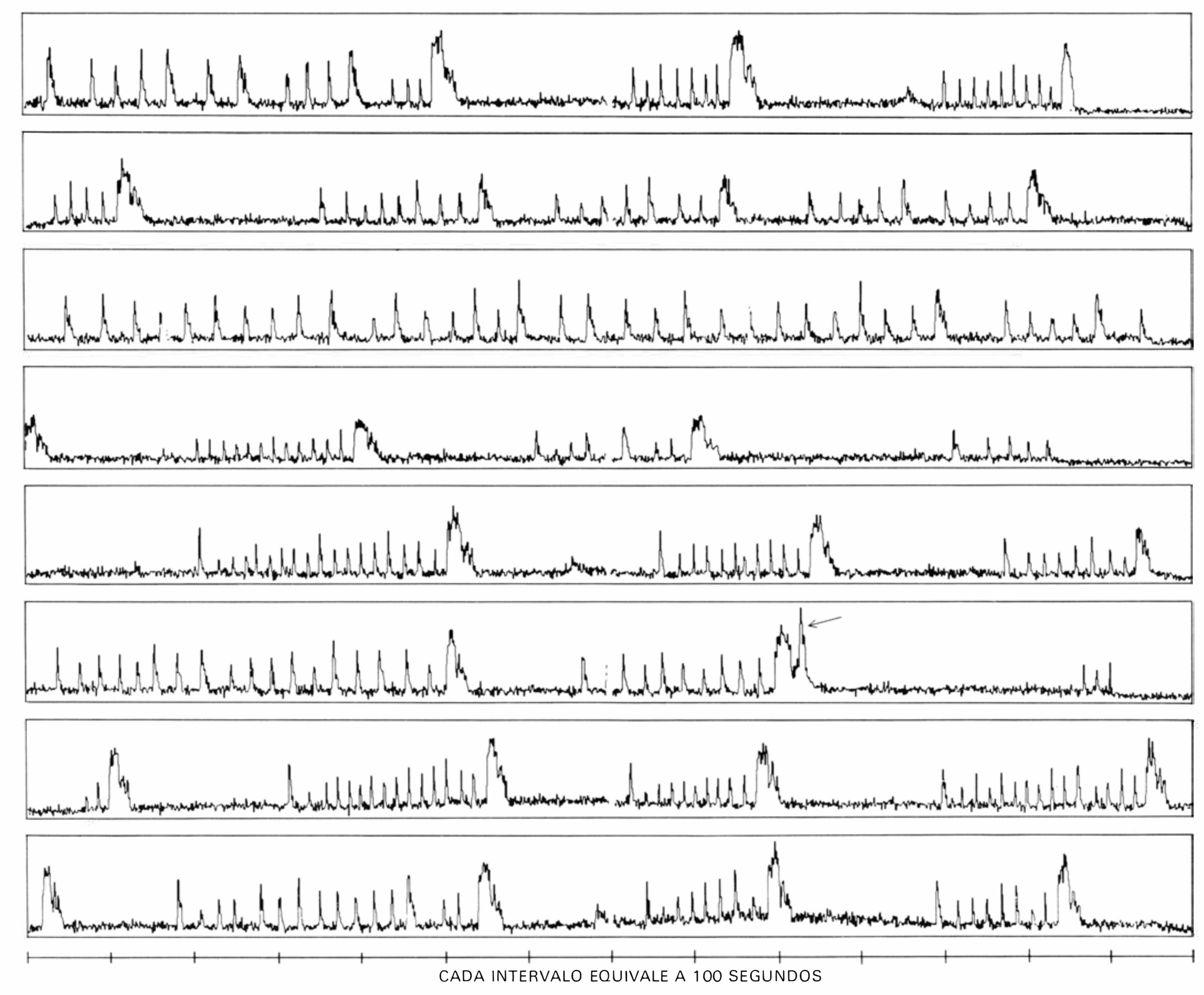
Todo esto ocurría en una reunión de la High Energy Astrophysics Division of the American Astronomical Society, en el M.I.T., en el mes de enero de 1976. Grindlay y Jernigan informaron de sus recientes resultados sobre las erupciones de NGC 6624. Richard Belian, Doyle Evans y Jerry Conner, del Los Alamos Scientific Laboratory, describieron sus observaciones, llevadas a cabo por los satélites Vela, de erupciones de rayos X a partir de una fuente sita en la constelación Norma; estos satélites eran los mismos que habían descubierto erupciones misteriosas de energías algo mayores (véase "Erupciones de rayos gamma cósmicos", por Ian B. Strong y Ray W. Klebesadel; INVESTIGACION Y CIENCIA, diciembre, 1976). Los investigadores norteamericanos desconocían entonces que los detectores de rayos X de altas energías del satélite ruso Kosmos 428 habían registrado dos breves aumentos del ritmo de cuenta durante una observación que duró algunas horas en 1971. En un artículo aparecido en la revista *Soviet Astronomy Letters*, publicado en el número correspondiente al mes de enero de 1975, los aumentos se atribuyeron a "erupciones de rayos X duros" a partir de dos fuentes cósmicas distintas. En nuestra sala de lectura apareció una traducción del artículo en marzo de 1976. Un análisis detallado de las propiedades de esas erupciones y de las circunstancias de su observación indican, sin embargo, que se originaron en la magnetosfera terrestre.

En las mismas fechas de las reuniones de la American Astronomical Society, el grupo del SAS-3 estaba llevando a cabo otra de las observaciones previstas de la fuente 3U 1820-30 en NGC 6624, desde la sala de operaciones situada en un edificio cercano, utilizando esta vez un conjunto de detectores distinto, con colimadores de haz fino y de haz abierto para definir el campo visual. Manteniendo el centro de esos campos visuales cerca del blanco, por medio del sistema giroscópico de control del satélite, que ajustamos de acuerdo con nuestros análisis de los datos de la cámara estelar, registramos una elevada intensidad de rayos X y algunas erupciones. Las erupciones eran desconcertantes,

debido a que no se registraron en todos los detectores, como uno hubiera esperado si todas hubieran venido de una sola fuente situada en la dirección (NGC 6624), hacia la cual apuntaba el campo visual de haz fino. En vez de ocurrir así aparecían en un detector con un colimador de haz abierto, cuyo estrecho campo visual pasaba sobre el centro de la galaxia. En la siguiente observación programada, Walter H. G. Levin y Jeffrey Hoffman, de nuestro grupo, dirigieron un equipo de investigadores del SAS-3 para el estudio de la estrella de rayos X pulsante y cercana GX1+4. También ellos observaron erupciones en un único detector de haz abierto, a partir de dos fuentes, por lo menos; de ellas, una podía estar cerca del centro galáctico y la otra no. Poco después, los detectores del SAS-3 se apuntaron directamente al centro de la galaxia, y Lewin y sus colegas descubrieron allí tres nuevos "eruptores", comprendidos entre los 0,2 grados a partir del centro.

Varias semanas más tarde, en búsqueda del cuarto nuevo "eruptor", Lewin apuntó los detectores hacia una posición situada a unos cinco grados del centro galáctico, en donde su análisis de los acontecimientos de la erupción anterior en un solo detector indicaba la presencia de otro eruptor. Habiéndonos acostumbrado, ya por entonces, a los eruptores lentos, con intervalos de repetición que van de horas a días, nos quedamos todos boquiabiertos al contemplar una secuencia de erupciones "de ametralladora", con intervalos de sólo 10 segundos. Se puso de manifiesto en seguida que las erupciones de mayor energía iban seguidas de intervalos más largos, y Saul A. Rappaport, de nuestro grupo, demostró muy pronto que la correlación era bastante lineal. A lo largo de los dos meses siguientes, Lewin y sus colegas observaron el comportamiento de este objeto único, MXB 1730-335; David Hearn midió su posición con una precisión de cinco minutos de arco. Poco después de empezar el intenso trabajo sobre el eruptor rápido, el otro eruptor, que inicialmente se había buscado en las proximidades, se localizó a 0,7 grados más lejos.

Pese a la notable variabilidad de sus características de recurrencia, el eruptor rápido demostró regularidades, que resultaban muy esclarecedoras de la naturaleza de su mecanismo. La relación aproximadamente lineal entre los eruptores y el intervalo que los sigue se parece a la que se espera que ocurra en un oscilador de amortiguamiento, como el



SECUENCIA DE ERUPCIONES de cadencia rápida, registrada a partir de MXB 1730-335, fuente de rayos X situada en un cúmulo globular desconocido anteriormente, que cae cerca del centro del disco galáctico. El intervalo entre las erupciones es, a veces, muy corto, de 10 segundos. El tiempo de recuperación después de cada erupción resulta proporcional, aproximadamente, al tamaño de éste. Sin embargo, la intensidad culminante, de una erupción a la siguiente, es casi la misma. Este comportamiento sugiere que la fuente se parece a un oscilador de amortiguamiento con un control, probablemente representado por la presión de radiación, que limita la amplitud. La flecha del sexto trazo indica una erupción a partir de un eruptor lento y cercano, el cual no se encuentra situado, en absoluto, en el interior de ningún cúmulo globular.

cebo de una lámpara de neón, que tiene un nivel de disparo constante y una profundidad de descarga variable. Después de cada disparo, el oscilador ha de recuperarse durante un tiempo, cuya duración es proporcional a la profundidad de la descarga, antes de poder dispararse de nuevo. Evidentemente, el eruptor rápido tiene un depósito de energía que se llena, a un ritmo más o menos estacionario, con la energía potencial gravitatoria de la materia que cae. Para cierto valor crítico de la masa acumulada, el depósito de energía "pierde" y se genera una erupción de rayos X. Otra pista del mecanismo del eruptor rápido es que todas sus erupciones, sea cual fuere su flujo total de energía, alcanzan, poco más o menos, el mismo máximo de intensidad. Sólo la duración de las erupciones varía en un factor considerable. Esto indica que el proceso de conversión de energía es autolimitado, tal vez a resultas del efecto de la presión de radiación y del calentamiento sobre el movimiento del plasma que cae por la gotera del depósito.

Tras el anuncio de la posición del eruptor rápido, William Liller, del Harvard College Observatory, tomó una placa de larga exposición, sensible al rojo, del campo estelar con el nuevo y potente telescopio de cuatro metros puesto recientemente en servicio en el Observatorio Interamericano de Cerro Tololo, en Chile. La placa puso de manifiesto un objeto nebuloso tenue dentro del círculo de error del eruptor rápido. Liller llegó a la conclusión de que el objeto es, probablemente, un cúmulo globular desconocido con anterioridad y muy oscurecido.

Observaciones subsiguientes, hechas en el infrarrojo por Douglas E. Kleinmann, del Harvard College Observatory, en colaboración con Susan G. Kleinmann y Edward Wright, del M.I.T., confirmaron que el objeto es, efectivamente, un cúmulo globular con una fuerte concentración de estrellas en su centro. La placa tomada por Liller no mostraba ninguna traza de cúmulo globular en la posición del eruptor lento cercano. Con todo, se sabía ya de dos cúmulos globulares que contenían fuentes eruptivas de rayos X. Los partidarios de la noción de que las fuentes de todas las erupciones eran agujeros negros masivos, situados en el núcleo de cúmulos globulares, empezaron a echar las campanas al vuelo. Incidentalmente, el eruptor rápido se apagó en abril de 1976 y no volvió a encenderse hasta 13 meses más tarde.

Emprendimos entonces otra observación del NGC 6624 con el SAS-3. Esta vez encontramos a 3U 1820-30 emitiendo erupciones, pero a partir de un estado de brillo reducido. En un período de cuatro días, observamos 22 erupciones, con intervalos entre ellas que empezaron siendo de unas 3,4 horas y disminuyeron gradualmente hasta 2,2 horas. (Como antes, nos perdimos algunas erupciones cuando la fuente era eclipsada por la tierra.) A lo largo de cuatro días, la persistente intensidad de 3U 1820-30 aumentó gradualmente hasta alcanzar más o menos un tercio de su valor brillante normal. Al llegar a ese punto, las erupciones se pararon, al parecer debido a algún cambio en las condiciones relacionadas con el aumento de la intensidad persistente.

La nueva observación demostró varios hechos clave. En primer lugar, la correlación entre aumento de brillo y cese de las erupciones suprimió cualquier duda que pudiera quedar acerca de que 3U 1820-30, situado en el centro del NGC 6624 o en su vecindad, es la fuente de las erupciones. En segundo lugar, el cambio en los intervalos entre erupciones demostró que el mecanismo de la erupción no es un buen reloj, como lo sería uno controlado por movimiento orbital o de rotación, sino que es un oscilador de amortiguamiento con una fuente de potencia variable. En cuanto a la naturaleza del misterioso glóbulo de la placa de NGC 6624 tomada por Neta Bahcall, una fotografía extraordinariamente clara, de corta exposición, de la región central hecha por Litter con el telescopio de cuatro metros de Cerro Tololo, muestra que se compone de supergigantes rojas, apiñadas en una región de un solo año-luz de diámetro. Habida cuenta de las estrellas tenues (tal vez 200 por cada supergigante roja), se llega a una estimación de la densidad central, que es de unas 3000 masas solares por año-luz cúbico.

Hacia junio de 1977, se habían localizado 29 fuentes repetitivas de erupciones, 17 de ellas con el SAS-3, nueve con el experimento del Goddard Space Flight Center de la NASA a bordo del OSO-8, y tres con los satélites ANS, *Uhuru* y Vela. De entre ellas, sólo cuatro parecen estar situadas en cúmulos globulares. Son varias las que claramente no están en cúmulos globulares. La conclusión, por consiguiente, es obvia: los eruptores, al igual que las fuentes de rayos X en general, ocurren de preferencia, pero no exclusivamente, en cúmulos globulares.

Resultados recientes, aportados por el satélite de rayos X británico Ariel 5, han venido a burlarse de la hipótesis de que sólo los cúmulos globulares centralmente condensados contienen estrellas de rayos X. Kenneth A. Pounds y sus colaboradores, de la Universidad de Leicester, descubrieron una estrella de rayos X asociada según parece al cúmulo globular, relativamente abierto, NGC 6712. Las mediciones del SAS-3 han confirmado la asociación, con una determinación de la posición que sitúa la fuente dentro de un minuto de arco del centro del cúmulo. Con ello, nuestras ideas aceptadas sobre la naturaleza de las estrellas de rayos X de cúmulos, que suponen que éstas surgen en núcleos densos de cúmulo, pueden verse en dificultades. Por otra parte, cabe la posibilidad de que estemos presenciando, en el NGC 6712, el efecto de batidora en trance de dispersar el cúmulo que había sufrido su fatal colapso central hace unos centenares de millones de años. No hay, sin embargo, prueba observacional de que un tal colapso dé lugar a un agujero negro masivo. Trabajos recientes de Stuart Shapiro, de Cornell, y de Alan P. Lightman, de Harvard, indican que el colapso se detendría y se invertiría tras la formación de un pequeño agujero negro (de 10 a 100 masas solares), o quizá la formación de varias binarias cerradas.

Hay buenas perspectivas para un rápido progreso en esta sección de la astronomía de altas energías. El 12 de agosto del presente año, se lanzó el primer observatorio astronómico de altas energías de la NASA, el HEAO-A. Equipado con instrumentos substancialmente más sensibles que los de los satélites anteriores, ampliará de inmediato la exploración de la emisión de rayos X, a partir de cúmulos globulares, a fuentes más de diez veces más tenues que los actuales límites de observación del *Uhuru* y del SAS-3.

El próximo año se lanzará el HEAO-B, que llevará el primer telescopio de rayos X con formación de imagen incorporado a bordo de un satélite. El instrumento tendrá una resolución angular de dos segundos de arco y supondrá un aumento de la sensibilidad en mil veces. Uno de los primeros objetivos de la misión será la medición sistemática de las posiciones de las fuentes de rayos X dentro de los cúmulos globulares. La desviación de las posiciones, con respecto al centro de los cúmulos, dará una medición estadística de la masa media de las estrellas de rayos X, y debería zanjar la cuestión de si son o no agujeros negros masivos. Además, la sensibilidad del HEAO-B permitirá explorar incluso la emisión de rayos X a partir de cúmulos globulares de otras galaxias. Los 100 cúmulos de la galaxia de Andrómeda serán blancos fáciles, y los 10.000 cúmulos globulares de galaxia elíptica gigante M87 también se investigarán.

Partículas fundamentales con encanto

La física de alta energía se ha ocupado con ahínco de la búsqueda de partículas con esta propiedad mecánico-cuántica. Se han descubierto varias de ellas en los residuos de las aniquilaciones electrón-positrón

Roy F. Schwitters

A nivel de las partículas elementales, las propiedades de la materia son notablemente escasas. Una partícula puede tener masa, o energía, y puede tener momento, incluyendo el momento angular intrínseco llamado espín. Puede tener carga eléctrica. Hay otras propiedades misteriosas, como la llamada extrañeza, pero no muchas más. En la mayoría de los casos, una lista de media docena de atributos describe completamente una partícula. Nada más puede decirse acerca de ella.

Debido a que hay tan pocas propiedades fundamentales de la materia, el descubrimiento de una nueva es un suceso importante en física. Recientemente, se ha hallado una de esas nuevas propiedades: tiene el caprichoso nombre de encanto ("charm"). Los átomos de la materia ordinaria no tienen encanto; esta propiedad puede observarse solamente en los residuos de las colisiones de alta energía entre partículas.

La primera señal de la existencia del encanto se registró en 1974, cuando se descubrieron partículas que tenían encanto en una forma escondida o latente. Con posterioridad, se han detectado partículas con encanto manifiesto. Estas nuevas partículas se encuentran, sin lugar a dudas, entre los descubrimientos más importantes de la física de altas energías durante la última década. Aun más, al esclarecer la cuestión del encanto, los físicos han aprendido mucho sobre la estructura de la materia ordinaria.

La mayoría de las partículas observadas en la naturaleza pertenecen a una u otra de estas dos clases: leptones y hadrones. Los leptones comprenden sólo cuatro partículas conocidas: el electrón, el muón y dos clases de neutrino. El electrón y el muón tienen ambos carga eléctrica -1, y son también esencialmente iguales en todas las demás propiedades, con excepción de la masa (el muón es unas 200 veces más pesado que el electrón). Los neutrinos carecen de carga eléctrica y, según parece, también de masa. Hay cuatro antileptones, idénticos a los correspondientes leptones en algunos aspectos, la masa por ejemplo, y diametralmente opuestos respecto a otras propiedades de los leptones. Verbigracia, el antielectrón o positrón tiene carga eléctrica + 1.

Los leptones se consideran elementales porque no pueden romperse en entidades más pequeñas. No tienen tamaño medible y no manifiestan ninguna estructura interna. Los hadrones, por el contrario, son objetos complejos y hay pruebas claras de que poseen una estructura interna. Se han identificado más de 100 clases de hadrones, de los cuales los más familiares son el protón y el neutrón, constituyentes del núcleo atómico. La gran variedad de hadrones y sus propiedades como partículas individuales los diferencian netamente de los leptones.

Una teoría, hoy ampliamente aceptada, explica verosímilmente tal diversidad. Lo hace estableciendo que los hadrones no son en absoluto partículas elementales sino entidades compuestas, formadas a su vez por unos pocos constituyentes más simples llamados quarks. En muchas de sus propiedades, los quarks resultan muy similares a los leptones; por ejemplo, deben ser partículas simples y puntuales. Pero no existe duda alguna en que los quarks se sitúan en una clase distinta de la clase de los leptones. Las interacciones entre los quarks están dominadas por una fuerza que no afecta en absoluto a los leptones.

Los físicos reconocen cuatro fuerzas básicas en la naturaleza; por orden de intensidad creciente, son: las fuerzas gravitacionales, las débiles, las electromagnéticas y las fuertes. La gravitación afecta a todas las partículas; su alcance es ilimitado, si bien su efecto en las partículas subatómicas es despreciable. La fuerza débil afecta también a todas las formas de materia. Aunque la fuerza débil es, en muchos órdenes de magnitud, más intensa que la gravitacional, se muestra aun lo suficientemente débil como para que se la pueda observar, a no ser que estén inhibidas las interacciones más fuertes.

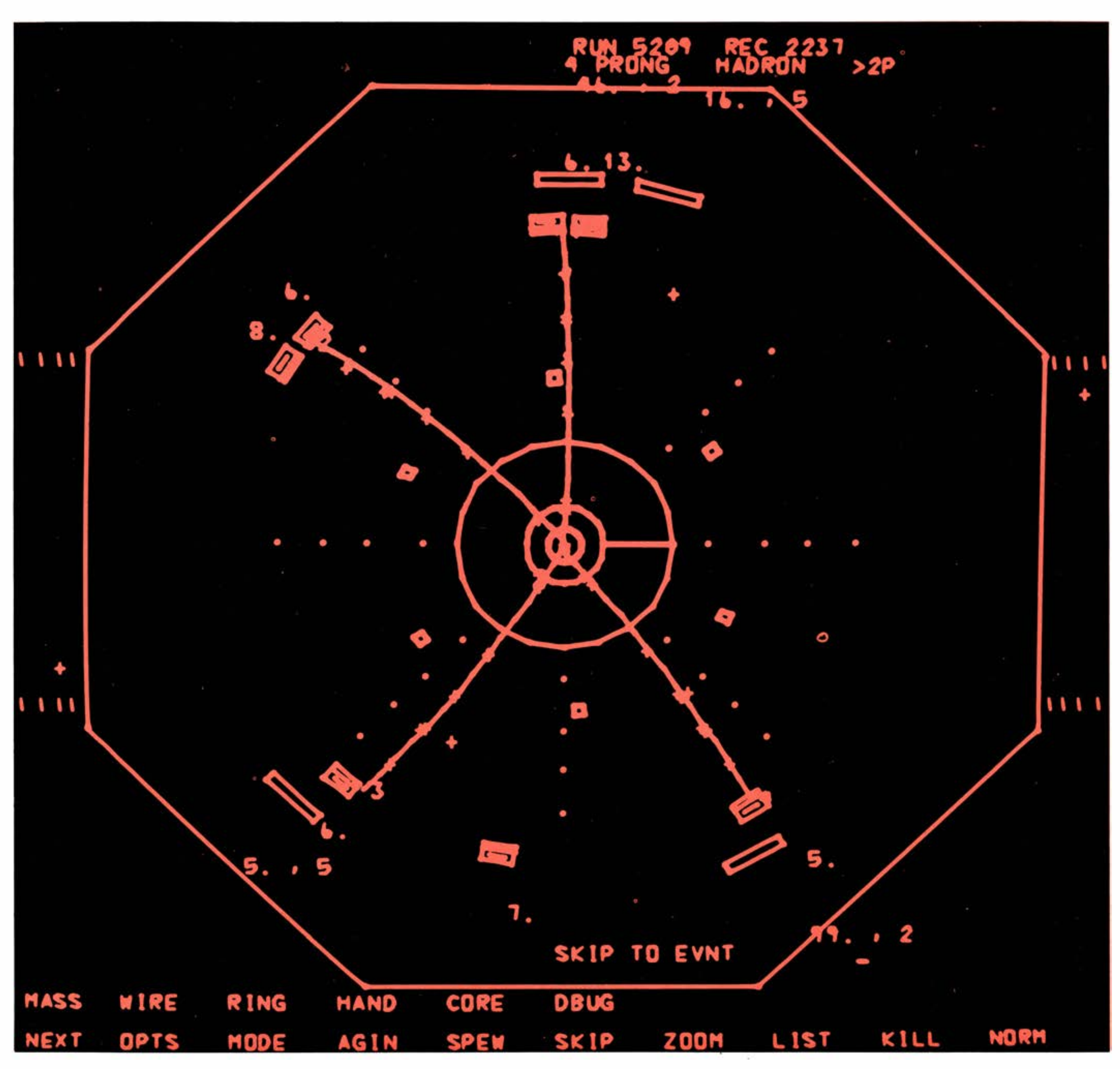
La fuerza electromagnética actúa exclusivamente sobre partículas que tienen carga eléctrica; entre estas partículas están el electrón y el muón y todos los quarks. Las fuerzas electromagnéticas enlazan los átomos entre sí y son responsables de casi todas las propiedades macroscópicas de la materia, incluidas las propiedades químicas.

La interacción fuerte distingue entre leptones y hadrones o, de acuerdo con la teoría, entre leptones y quarks. Todos los leptones son absolutamente insensibles a la interacción fuerte; sólo los quarks y los hadrones (que se suponen hechos de quarks) sienten su influencia. Los quarks pueden interaccionar con los leptones mediante las fuerzas electromagnéticas y débiles, pero entre sí mismos interaccionan casi exclusivamente mediante la interacción fuerte. La intensidad de esta fuerza supera en más de 100 veces a la de la fuerza electromagnética; a las energías estudiadas hoy día, resulta cerca de 10^{10} veces más intensa que la interacción débil.

Una teoría que da cuenta de todas las variedades de la materia con sólo unos pocos quarks y leptones representa una atrayente economía, pero implica

una renuncia significativa. Aunque la teoría ha ganado amplia aceptación, no se dispone de ninguna prueba de que los quarks existan aislados. Hasta el presente, nadie ha podido extraer un quark de un hadrón. Algunos teóricos sugieren que los quarks pueden estar tan fuertemente entrelazados dentro de los hadrones que nunca podrán aislarse en el laboratorio. La cuestión de la existencia de los quarks no se analizará aquí; los consideraremos como un medio para interpretar las relaciones entre partículas observadas experimentalmente.

La hipótesis de los quarks fue propuesta en 1963 independientemente por Murray Gell-Mann y George Zweig, pertenecientes ambos al Instituto de Tecnología de California. En la versión original del concepto había tres clases de quarks, denominados *u, d* y *s* (del inglés "up" –arriba–, "down" –abajo– y "strange" –extraño–; y estaban los tres antiquarks correspondientes, denomina-



DESINTEGRACION de una partícula encantada, reconstruida en una imagen generada por un ordenador sobre un tubo de rayos catódicos. La imagen incluye una sección transversal y esquemática del detector de partículas. El octágono es la forma básica del detector; las líneas de cuatro puntos indican las posiciones radiales de cuatro cámaras de chispas cilíndricas; el círculo más interior, en el centro de la imagen, representa el tubo del haz en el que los electrones y sus antipartículas, los positrones, chocan para producir otras partículas. Las trayectorias de las partículas se obtuvieron a partir del disparo de descargas en la cámara de chispas y de dos anillos concéntricos de contadores de centelleo, de los que los más internos están representados por pequeños cuadrados y los externos por rectángulos. En este suceso, la trayectoria de la partícula en la posición de las 12 en punto se ha identificado como la de un mesón *K* cargado negativamente (K^-) y la trayectoria en la posición de las siete en punto es la de un pión positivo (π^+). Se supone que estas dos partículas son los productos de desintegración de una partícula, llamada D^o, que lleva la nueva propiedad de la materia llamada encanto. La D^o se desintegra demasiado de prisa como para poder detectarse directamente. En las colisiones electrón-positrón, las partículas encantadas pueden crearse sólo en combinación con antipartículas encantadas; en este caso, se cree que el compañero de antimateria de la D^o es un estado excitado, el $\bar{D}^{*o}$, cuyos productos de desintegración se han identificado sólo de una manera provisional. Los contadores de centelleo, cerca de la parte inferior de la imagen, que no tienen asociada ninguna trayectoria de partícula, pueden haberse disparado por un rayo gamma emitido cuando el D^{*o} se desintegraba en un D^o. Esta partícula podía haberse desintegrado dando un muón (μ^-) en la posición de las 10 en punto, un mesón *K* (K^+) en la posición de las cinco en punto y un neutrino que no se habría detectado.

LEPTONES	MASA (GeV)	ESPIN	CARGA ELECTRICA
NEUTRINO ELECTRON ν_e	0	½	0
ELECTRON e^-	.0005	½	−1
NEUTRINO MUONICO ν_μ	0	½	0
MUON μ^-	.105	½	−1

QUARKS	MASA (GeV)	ESPIN	CARGA ELECTRICA	EXTRAÑEZA	ENCANTO
u	.1	½	+⅔	0	0
d	.1	½	−⅓	0	0
s	.4	½	−⅓	−1	0
c	1.5	½	+⅔	0	+1

LEPTONES Y QUARKS son las únicas clases importantes de partículas que hoy se consideran elementales. Según parece, ambos son entidades simples y puntuales, sin estructura interna y sin tamaño medible. Hay cuatro leptones conocidos, agrupados en parejas. En la formulación inicial de la teoría de los quarks, había sólo tres quarks: los llamados *u* y *d* que formaban una pareja, y el quark *s*, que carecía de compañero. La hipótesis del encanto establece una simetría entre leptones y quarks añadiendo un cuarto quark, llamado *c*. Debe existir una relación entre el quark *s*, que comporta la propiedad de la materia denominada extrañeza, y el quark *c*, que tiene la propiedad análoga del encanto. Para cada leptón y quark existe una antipartícula de propiedades exactamente opuestas. No se ha logrado aislar a los quarks, pero, al parecer, son los constituyentes de otras partículas.

	HADRON	SIMBOLO	COMPOSICION EN QUARKS	MASA (GeV)	ESPIN	VIDA MEDIA (SEG.)	CARGA ELECTRICA	EXTRAÑEZA	ENCANTO
BARIONES	PROTON	p	uud	.938	½	ESTABLE	+1	0	0
	NEUTRON	n	udd	.940	½	10^3	0	0	0
	LAMBDA	Λ	uds	1.116	½	10^{-10}	0	−1	0
	LAMBDA ENCANTADA	Λ_c	udc	2.260	½	?	0	0	+1
MESONES	MESONES PI	π^+	$u\bar{d}$	.140	0	10^{-8}	+1	0	0
		π^-	$d\bar{u}$	.140	0	10^{-8}	−1	0	0
		π^0	$u\bar{u} + d\bar{d}$	.135	0	10^{-16}	0	0	0
	MESONES K	K^+	$u\bar{s}$	.494	0	10^{-8}	+1	+1	0
		K^-	$s\bar{u}$	.494	0	10^{-8}	−1	−1	0
	PHI	ϕ	$s\bar{s}$	1.020	1	10^{-22}	0	0	0
	PSI	ψ	$c\bar{c}$	3.095	1	10^{-20}	0	0	0
	MESONES ENCANTADOS	D^0	$c\bar{u}$	1.863	0	?	0	0	+1
		D^+	$c\bar{d}$	1.868	0	?	+1	0	+1
		F^+	$c\bar{s}$	?	0	?	+1	+1	+1

LOS HADRONES son las partículas formadas por quarks. Los llamados mesones constan de un quark y un antiquark; este último se representa por el símbolo del quark con una barra sobre él. Los bariones constan de tres quarks ligados; hay también antibariones que combinan tres antiquarks. Las propiedades de los hadrones (y de otras partículas) se describen por medio de enteros y fracciones sencillas llamados números cuánticos. El primer hadrón que hizo sospechar la existencia del encanto fue el mesón psi, con la estructura de quarks $c\bar{c}$, pero como los números cuánticos de encanto de estos quark y antiquark se cancelan, el encanto total de la psi es nulo. Los hadrones de menor masa que presentan encanto manifiesto son los mesones D^0, D^+ y F^+, y se han podido observar ya los dos primeros. Se ha sugerido la existencia de un barión encantado, denominado Λ_c.

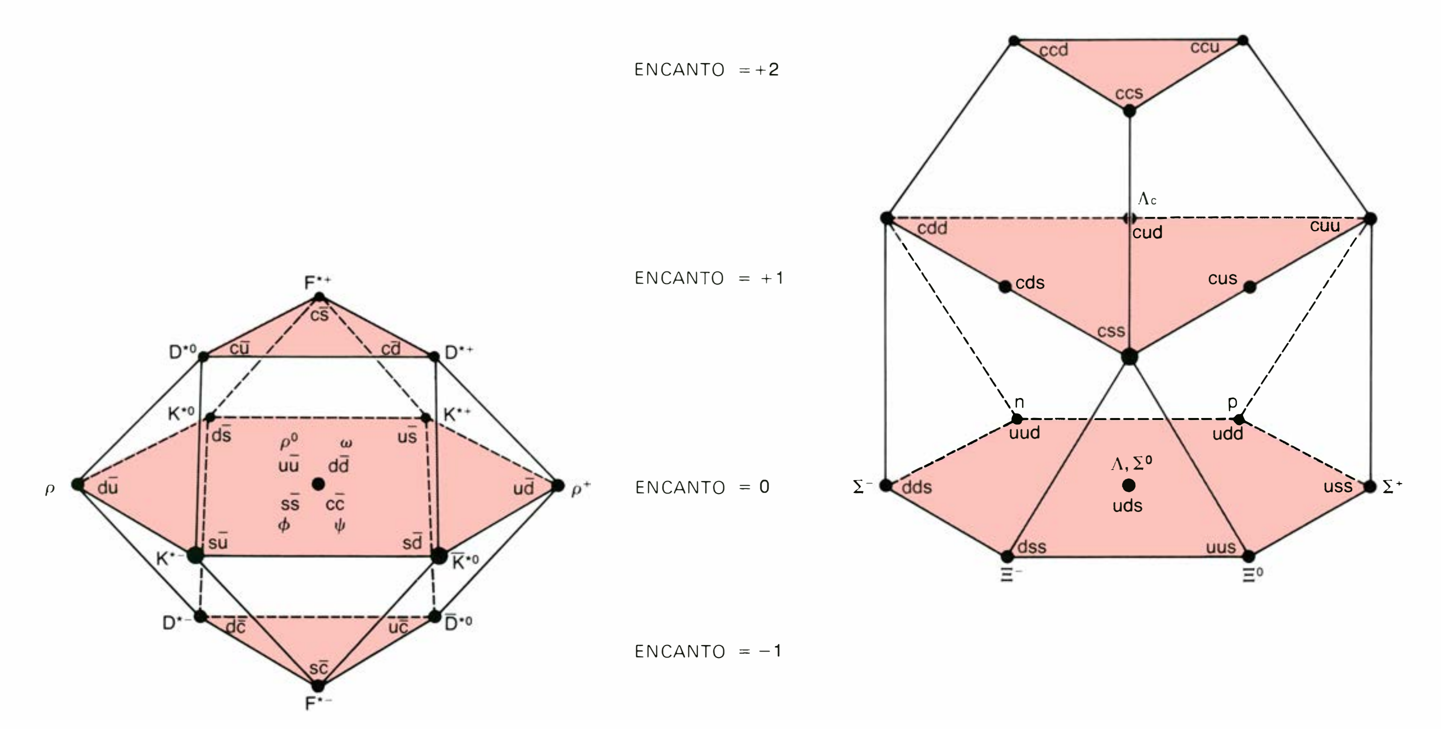
dos $\bar{u}$, $\bar{d}$ y $\bar{s}$. Los hadrones se forman combinando los quarks y los antiquarks según reglas sencillas. Una posibilidad es ligar un quark y un antiquark; la partícula resultante es miembro de la clase de hadrones llamados mesones; ejemplo de ello es el pión cargado positivamente (π^+), que está formado por un quark *u* y un antiquark $\bar{d}$. Otra combinación permitida consiste en un estado ligado de tres quarks. Los hadrones formados de esta manera se llaman bariones; comprenden el protón (con la composición de quarks *uud*) y el neutrón *(udd)*. Finalmente, los antibariones pueden formarse por combinaciones de tres antiquarks.

Estas son las únicas maneras permitidas de combinar los quarks para formar hadrones. Otras combinaciones, fácilmente imaginables, serían las partículas construidas a partir de dos quarks o de un quark y dos antiquarks, pero tales hadrones no existen.

Las propiedades observadas de los hadrones se explican de manera directa a partir de las propiedades asignadas a sus quarks constituyentes. Con la excepción de la masa, todas las propiedades de la materia que son necesarias para identificar a cualquier partícula elemental aparecen sólo en unidades discretas, o cuantos; pueden medirse, por tanto, mediante enteros o fracciones sencillas llamados números cuánticos. El momento angular de espín, por ejemplo, se observa sólo en cantidades enteras o semienteras (cuando se mide en unidades naturales); no son posibles valores intermedios. En todas las partículas observadas, la carga eléctrica se presenta sólo en unidades enteras de la carga del electrón.

La mayoría de los números cuánticos de un hadrón vienen determinados por simple adición de los números cuánticos de los quarks constituyentes. En el caso de la carga eléctrica, este procedimiento exige que los quarks tengan asignados números cuánticos un tanto raros: deben poseer carga eléctrica fraccionaria. El quark *u* tiene una carga +2/3; los quarks *d* y *s* tienen ambos una carga −1/3. Las cargas de los antiquarks son las opuestas. Así, en el pión cargado positivamente, formado por un quark *u* y un antiquark $\bar{d}$, las cargas +2/3 y +1/3 suman +1. Los quarks *uud* del protón, con cargas +2/3, +2/3 y −1/3, también dan una suma de +1. En el neutrón *(udd)*, las cargas +2/3, −1/3 y −1/3 suman cero. Todos estos resultados están de acuerdo con las cargas eléctricas conocidas de los hadrones.

Otros números cuánticos pueden tratarse de manera similar. La extrañeza,



LAS FAMILIAS DE HADRONES definen todas las combinaciones de quarks permitidas. Hay familias distintas para cada posible valor del momento angular intrínseco de espín; las mostradas aquí son los mesones con una unidad de espín y los bariones con media unidad de espín. Hasta hace poco, sólo se había observado las partículas situadas en el plano marcado "encanto = 0", pero el descubrimiento de partículas encantadas implica que todos los otros deben también existir. La partícula psi completa un grupo de cuatro partículas en el centro del conjunto de mesones de espín 1, Los D^o, D^+ F^+ son miembros de otra familia de mesones, los de espín cero, pero están representados en el grupo de espín 1 por los estados excitados simbolizados por D^{*o}, D^{*+} y F^{*+}. Se han encontrado, en el detector, el D^{*o}, el D^{*+} y sus antipartículas correspondientes.

por ejemplo, es un número cuántico asignado a ciertos hadrones cuya vida media es anormalmente larga. En el modelo de quarks, estas partículas se distinguen por la presencia de un quark *s* o un antiquark $\bar{s}$ que tienen números cuánticos de extrañeza −1 y + 1, respectivamente; los otros quarks tienen extrañeza cero. La extrañeza del hadrón se determina sumando los números cuánticos de extrañeza de todos los quarks constituyentes. El mesón *K* cargado positivamente (K^+), formado por un quark *u* y un antiquark $\bar{s}$, tiene extrañeza + 1.

El concepto de extrañeza se introdujo como explicación de la lenta desintegración de ciertos hadrones descubiertos en los años 1950. Los hadrones pesados suelen desintegrarse mediante la interacción fuerte, que actúa muy rápidamente; en cierta manera, la interacción se llama fuerte por ser tan rápida. Un hadrón que se desintegra fuertemente existe sólo durante 10^{-23} segundos, tras los cuales se rompe en hadrones menos pesados. Todas las propiedades de la materia deben permanecer inalteradas durante este proceso; por ejemplo, la carga eléctrica total de los productos de la desintegración debe ser igual a la carga eléctrica de la partícula que se desintegra. Esta exigencia se expresa diciendo que todos los números cuánticos deben conservarse.

Los hadrones extraños pueden tomar parte en tales desintegraciones siempre que el número cuántico de extrañeza pueda conservarse. En términos del modelo de los quarks, la necesidad de conservar la extrañeza implica que el quark extraño *s* debe pasar intacto a los productos de la desintegración. Las partículas extrañas más pesadas se desintegran así en partículas extrañas más ligeras y sus vidas medias no son sustancialmente más largas que las de los otros hadrones que se desintegran fuertemente. Pero, en último término, debe haber un conjunto de partículas extrañas más ligeras, que no tengan estados de menor masa a los cuales puedan dar el quark *s*. Estos son los mesones *K* y el barión lambda (Λ).

Las partículas *K* y lambda no pueden desintegrarse mediante la interacción fuerte. En realidad no podrían desintegrarse de ninguna manera si no fuese por una importante característica de las interacciones débiles, a saber: algunas leyes de conservación que se respetan estrictamente en todas las interacciones fuertes y electromagnéticas pueden ignorarse en las interacciones débiles. La extrañeza está entre estos números cuánticos más o menos conservados; en consecuencia, las interacciones débiles pueden convertir un quark *s* en un quark *u* o *d* y las partículas extrañas pueden desintegrarse mediante interacciones débiles en hadrones no extraños más ligeros o en leptones. Sin embargo, las partículas que sólo pueden desintegrarse mediante la interacción débil tienen vidas medias mucho más largas que los hadrones, que se desintegran fuertemente; las partículas *K* y lambda tienen vidas medias que van de 10^{-10} segundos a 10^{-8} segundos.

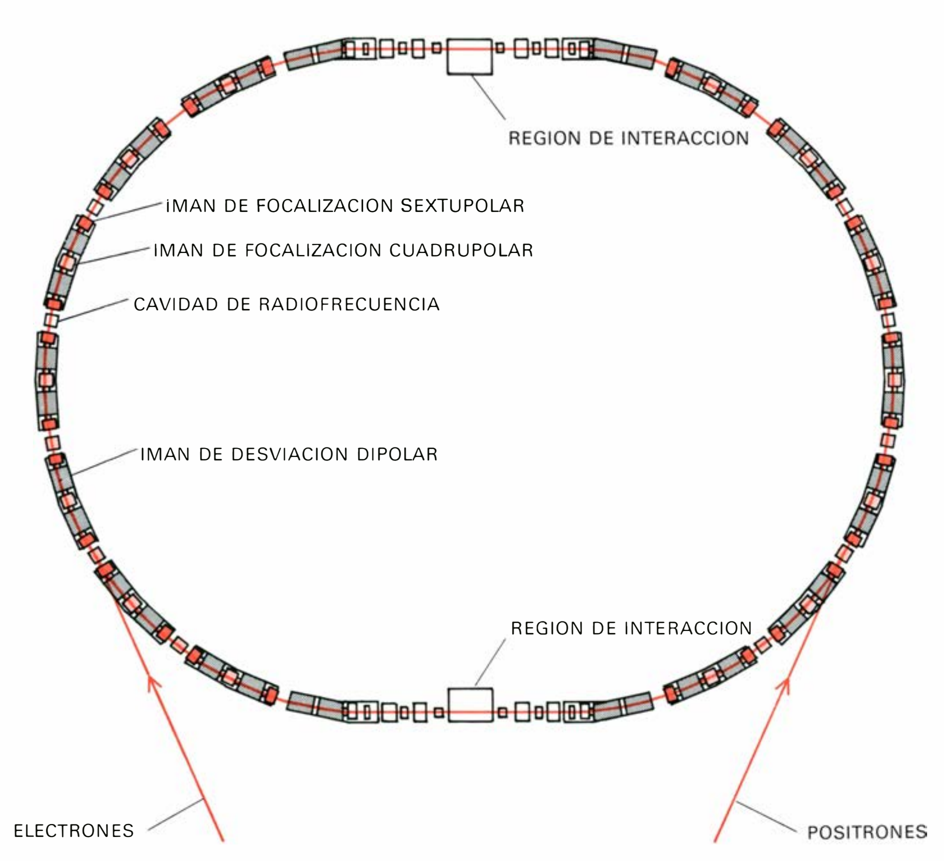
Hay unos números cuánticos que requieren un sistema de recuento ligeramente más elaborado que los de carga eléctrica y extrañeza. El más característico entre ellos es el momento angular de espín. Se supone que todos los quarks y leptones poseen la misma cantidad de espín, a saber: 1/2 unidad; las reglas de la mecánica cuántica exigen que tales partículas puedan tener su espín orientado en sólo dos direcciones posibles (con respecto a cualquier sistema de referencia arbitrario). Así, el quark y el antiquark de un mesón pueden tener sus espines en la misma dirección o en direcciones opuestas. En el primer caso, el momento angular resultante del mesón es 1; en el segundo es 0. De estas reglas se deduce que los mesones tienen siempre espín entero. En un barión hay también dos posibles configuraciones del espín: los espines de los tres quarks pueden ser paralelos o el espín de uno de

ellos puede estar dirigido en la dirección opuesta. Los correspondientes espines del barión son 3/2 y 1/2; los espines de los bariones son invariablemente semienteros. Complicando aun más la situación, un sistema ligado de quarks puede tener momento angular orbital además del espín: los quarks pueden girar unos en torno a otros o en torno a su común centro de masas, además de girar individualmente sobre sus ejes. El momento angular orbital añade incrementos enteros al momento angular total del hadrón. Debido a que hay muchos estados de espín posibles, una única combinación de quarks puede dar lugar a numerosas partículas con varias cantidades de momento angular y también con diferentes energías o masas. Son estados distintos de materia a pesar de que están formados por los mismos quarks.

Observando las reglas de combinación de los quarks y del contaje de sus números cuánticos, se pueden explicar todas las propiedades de los hadrones. Cada hadrón conocido puede definirse como una combinación de un quark y un antiquark, o bien de tres quarks. Aun más, a cada combinación permitida de quarks le corresponde un hadrón conocido. No hay vacantes.

Tal esquema de clasificación de los hadrones quedó en entredicho con el descubrimiento de una nueva partícula en 1974. La partícula era un hadrón, pero no podía formarse mediante ninguna combinación permitida de tres quarks, pues todas ellas ya tenían candidato.

La nueva partícula fue descubierta independientemente, y casi al mismo tiempo, por dos grupos experimentales que seguían técnicas muy distintas. Un grupo, cuyos miembros pertenecían al Instituto de Tecnología de Massachusetts y al Brookhaven National Laboratory, encontró la partícula en el transcurso de un experimento realizado en Brookhaven y la llamaron *J*. El otro grupo, en el que el autor estaba encuadrado, reunía físicos del Stanford Linear Accelerator Center (SLAC) y del Lawrence Berkeley Laboratory. Obtuvimos la prueba de la existencia de la nueva partícula en un experimento en el SLAC, y la bautizamos con la letra griega psi (ψ). Aquí yo adoptaré el último nombre. El año pasado, los directores de ambos grupos, Samuel C. C. Ting, del MIT, y Burton Richter, del SLAC, compartieron el premio Nobel por su descubrimiento.



EL ANILLO DE ALMACENAMIENTO DE PARTICULAS del Stanford Linear Accelerator Center (SLAC) fue uno de los dispositivos que permitieron el descubrimiento de la partícula psi; más recientemente, se han observado en él los mesones encantados. El anillo oval, llamado SPEAR, recibe electrones y positrones suministrados por el acelerador de 3219 metros del SLAC. Los electrones y positrones, que giran en sentido contrario, están confinados en el anillo mediante imanes que curvan las trayectorias de las partículas y focalizan los haces en "paquetes" en forma de aguja. Cada paquete consta de unas 10^{11} partículas, y los paquetes están sincronizados de manera que se cruzan en dos secciones transversales del anillo. Las pérdidas de energía de las partículas están compensadas por cavidades de radiofrecuencia. Los imanes obligan a que las partículas se desplacen en direcciones opuestas, toda vez que los electrones y positrones tienen cargas opuestas.

La nueva partícula tiene una masa de cerca de 3100 millones de electrónvoltios (3,1 GeV). Esto es, más de tres veces la masa del protón; ello determina que la psi sea una de las partículas conocidas más pesadas. Aquellos de sus números cuánticos que pudieron medirse han resultado ser completamente convencionales. La psi tiene momento angular de espín 1, la unidad entera que indica que la partícula debe ser un mesón. Es neutra eléctricamente y tiene extrañeza cero. Los productos de su desintegración son partículas familiares como piones, electrones y muones. Con estas propiedades, la psi hubiera podido pasar como un hadrón ordinario, excepcional sólo por su masa. El problema era que todos los estados de la materia con los números cuánticos de la psi habían sido identificados desde hacía tiempo con otros hadrones. En un mundo hecho de tres quarks no había necesidad de la psi; ni siquiera había lugar para ella.

Desde el punto de vista de la experimentación, el rasgo distintivo de la psi es su vida media, excepcionalmente larga. Se conocen muchos hadrones con masas grandes, pero casi todos ellos se desintegran muy rápidamente mediante las interacciones fuertes; sus vidas medias suelen ser del orden de 10^{-23} segundos. La partícula psi también se desintegra mediante la interacción fuerte, pero tiene una vida media de unos 10^{-20} segundos, mil veces más larga que la vida media de un hadrón típico de masa comparable. Se exige alguna explicación a esa longevidad.

Un precedente sugestivo de la larga vida media de la psi se encuentra en otra partícula, descubierta unos años antes: el mesón phi (φ). El mesón phi consta de un quark extraño ligado a un antiquark extraño; como los números cuánticos de extrañeza del quark y el antiquark se cancelan, ese mesón tiene extrañeza resultante cero. La phi puede desintegrarse por la interacción fuerte; sólo se requiere que el quark y el antiquark se aniquilen entre sí. Este proceso conserva todos los números cuánticos, pero a pesar de ello parece estar inhibido. La vida media de la phi es aproximadamente 10 veces más larga de lo que podría esperarse.

La manera más sencilla de ofrecer a la partícula psi un lugar entre los hadrones es considerarla un mesón constituido por un quark nuevo y pesado, ligado a su correspondiente antiquark. Esta estructura da cuenta automáticamente de los números cuánticos de la psi. La larga vida media puede explicarse por un mecanismo similar al que retarda la desintegración del mesón phi: las desintegraciones fuertes estarían permitidas, pero podrían retardarse de alguna manera.

Desde el momento de su descubrimiento, la explicación más extendida acerca de la partícula psi fue la de un estado ligado de un nuevo quark y un nuevo antiquark. Esta hipótesis debió, en gran parte, su credibilidad a la existencia de un cuarto quark, que se había propuesto mucho antes a partir de hechos que no tenían nada que ver con el descubrimiento de nuevos hadrones.

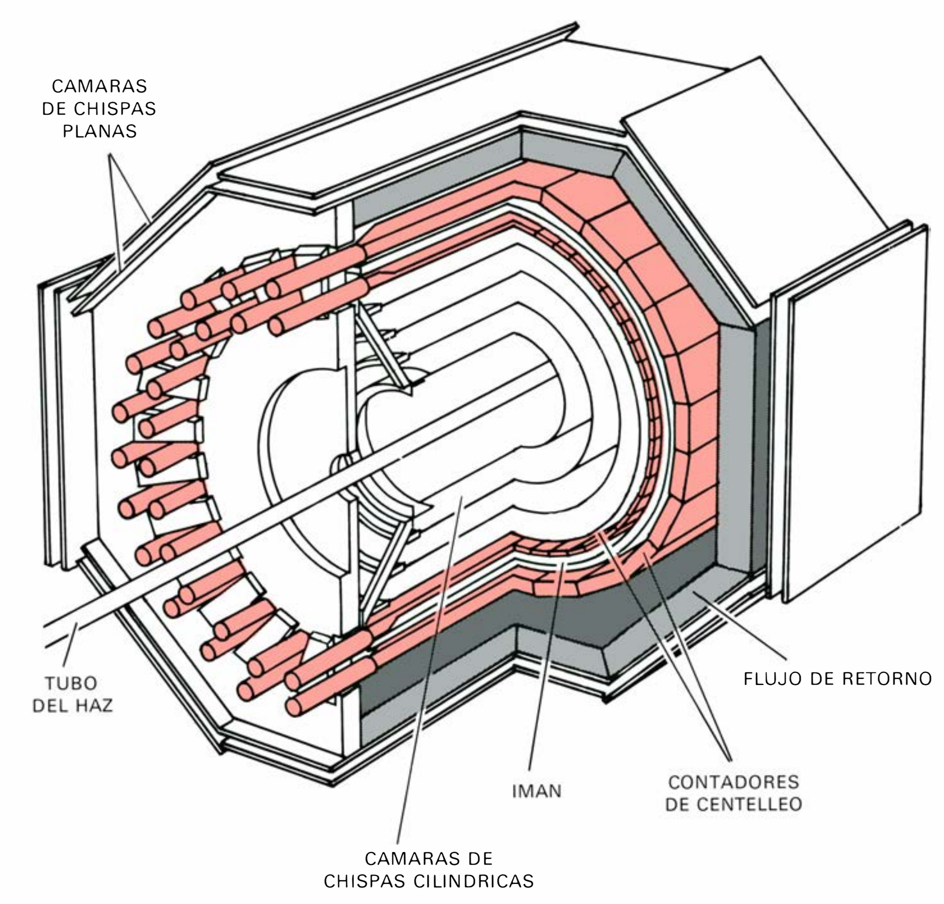
El nuevo quark fue sugerido por ciertos teóricos como una extensión natural del modelo formulado por Gell-Mann y Zweig. Inicialmente, estuvo basado en un argumento estético fundamentado en la idea de una profunda y posible conexión entre los leptones y los quarks. Como hay cuatro leptones conocidos, se argüía, el espectro de las partículas elementales sería mucho más atractivo si hubiese también cuatro quarks. Los leptones van emparejados; así, el electrón está asociado a un neutrino y el muón al otro. Los quarks u y d forman una pareja similar, pero el quark s no tiene compañero. El nuevo quark estaba destinado a llenar esta ausencia. Para hacerlo, debía tener una carga eléctrica $+2/3$ y algún nuevo número cuántico que lo distinguiera de los quarks conocidos. James D. Bjorken, del SLAC, y Sheldon Lee Glashow, de la Harvard University, dieron al número cuántico el nombre de "charm" (encanto). El quark encantado se designa por c y el antiquark encantado $\bar{c}$.

El encanto es una propiedad muy parecida a la extrañeza: debe conservarse en todas las interacciones fuertes y electromagnéticas pero no en las débiles. Por tanto, las partículas encantadas más ligeras deberían desintegrarse sólo vía la interacción débil y tener una vida media sensiblemente mayor. La partícula psi, sin embargo, escapa a esta regla porque su encanto total es cero. Consta de un quark c y un antiquark $\bar{c}$ cuyos números cuánticos de encanto se cancelan.

En 1970, Glashow, en colaboración con John Iliopoulos y Luciano Maiani, propuso una descripción teórica esencial para el quark encantado. Se fundaba en observaciones no relacionadas directamente con las partículas encantadas, pero sí con ciertas interacciones débiles de partículas extrañas. Las interacciones débiles pueden tener lugar con o sin transferencia de carga eléctrica entre las partículas interactuantes. Hacia 1970 se había establecido experimentalmente una correlación paradójica: con muy raras excepciones, en las interacciones débiles se cambiaba la extrañeza sólo cuando se transfería carga eléctrica. Las llamadas corrientes débiles neutras, que no transferían carga, no parecían variar tampoco la extrañeza.

En el modelo de los tres quarks no existía una conexión obvia entre la transferencia de carga y el cambio de extrañeza. Glashow, Iliopoulos y Maiani observaron que la adición de un cuarto quark, c, estrechamente asociado al quark s, podía explicar la supresión. El cómo ocurre esto se razona mediante un argumento bastante sutil de mecánica cuántica. El cuarto quark no impide directamente las interacciones que cambian la extrañeza sin transferir carga. Por el contrario, proporciona un canal alternativo para esas interacciones, aunque de suerte tal que los efectos de los dos canales se cancelan.

Cuando se descubrió la psi, las partículas encantadas eran ya piezas importantes en una larga lista de entidades cuya existencia se había avanzado teóricamente, pero no se habían observado nunca. La hipótesis del encanto podía explicar las propiedades de la psi, pero también podían hacerlo otras teorías. Por otro lado, el modelo del encanto prevé mucho más que la existencia de una única nueva partícula: si el quark encantado existe, debe haber un espectro completo de nuevos estados de materia. Algunos de estos estados están estrechamente relacionados con la psi: constan de un quark c y un antiquark $\bar{c}$, pero tienen masas diferentes y momento angular con valores distintos. El quark encantado debe combinarse también con los tres quarks originales para formar docenas de nuevos hadrones con encanto manifiesto. Mediante la búsqueda de estas nuevas partículas podríamos determinar si la partícula psi contiene realmente una nueva clase de quark y si



EL DETECTOR MAGNETICO del SPEAR desempeñó un papel crucial en el descubrimiento de la partícula psi y de los mesones encantados. El detector comprende dos clases de aparatos: cámaras de chispas y contadores de centelleo. Ambos son sensibles a partículas cargadas eléctricamente, emitidas en colisiones electrón-positrón en el centro del detector. Un campo magnético fuerte curva las trayectorias de las partículas cargadas; a partir de la cantidad de curvatura puede determinarse el momento. Los contadores de centelleo están hechos de un plástico que emite un destello de luz cuando una partícula lo atraviesa; miden el tiempo de vuelo de la partícula y, por tanto, su velocidad. Sabiendo el momento y la velocidad de una partícula, puede estimarse su masa en reposo.

dicho quark posee o no el número cuántico de encanto propuesto.

En el SLAC, el experimento en que se descubrió la psi y los que se realizaron subsiguientemente buscando otras partículas que poseyeran quarks encantados, se llevaron a cabo usando la misma técnica fundamental. Los electrones y sus antipartículas, los positrones, se hacían chocar a alta energía. Las partículas de interés se encuentran entonces buscando entre los residuos de las colisiones.

Las colisiones tienen lugar en un dispositivo llamado anillo de almacenamiento, en el que los electrones y positrones circulan en direcciones opuestas con velocidad cercana a la de la luz. El anillo de almacenamiento del SLAC se llama SPEAR; se construyó en 1971 y 1972 bajo la dirección de Burton Richter y John Rees. En el centro del anillo SPEAR hay una cámara de vacío toroidal de aluminio de un diámetro medio de 80 metros y una sección de varias decenas de milímetros. Cuando el anillo está en funcionamiento, unos 10^{11} electrones circulan en el sentido de las agujas del reloj formando un "paquete" en aguja, de unos pocos centímetros de largo y menos de medio milímetro de ancho. En dirección contraria circula un paquete similar de positrones. Los paquetes se cruzan dos veces durante cada vuelta y sus órbitas están programadas de manera que las colisiones ocurran en dos cortas secciones transversales del anillo, en donde se colocan detectores de partículas. En el resto de la circunferencia del anillo, la cámara de vacío está inmersa en grandes electroimanes, que curvan las trayectorias de las partículas para formar la vía circular y mantienen focalizados los paquetes. Hay también cuatro cavidades en las que se proporciona energía de radiofrecuencia para compensar la energía perdida por los electrones y positrones circulantes.

Los electrones y positrones se generan en el acelerador lineal de 3219 metros de longitud del SLAC. En un extremo del acelerador, se emiten los electrones por un filamento caliente, de manera muy similar a como lo hacen en un tubo de vacío ordinario. Los electrones se inyectan en el SPEAR acelerándolos y dirigiéndolos hacia el anillo. El suministro de positrones exige un procedimiento ligeramente más elaborado. Primero, los electrones avanzan hasta un tercio de la longitud del acelerador, donde chocan contra un blanco de cobre, produciendo un chorro de partículas secundarias, entre las que se encuentran electrones, positrones y rayos gamma. El resto de la longitud del acelerador se ajusta entonces de manera que acelere sólo los positrones. El llenado de los haces almacenados se produce en pocos minutos; transcurridos éstos, los haces circulan establemente durante períodos de hasta varias horas.

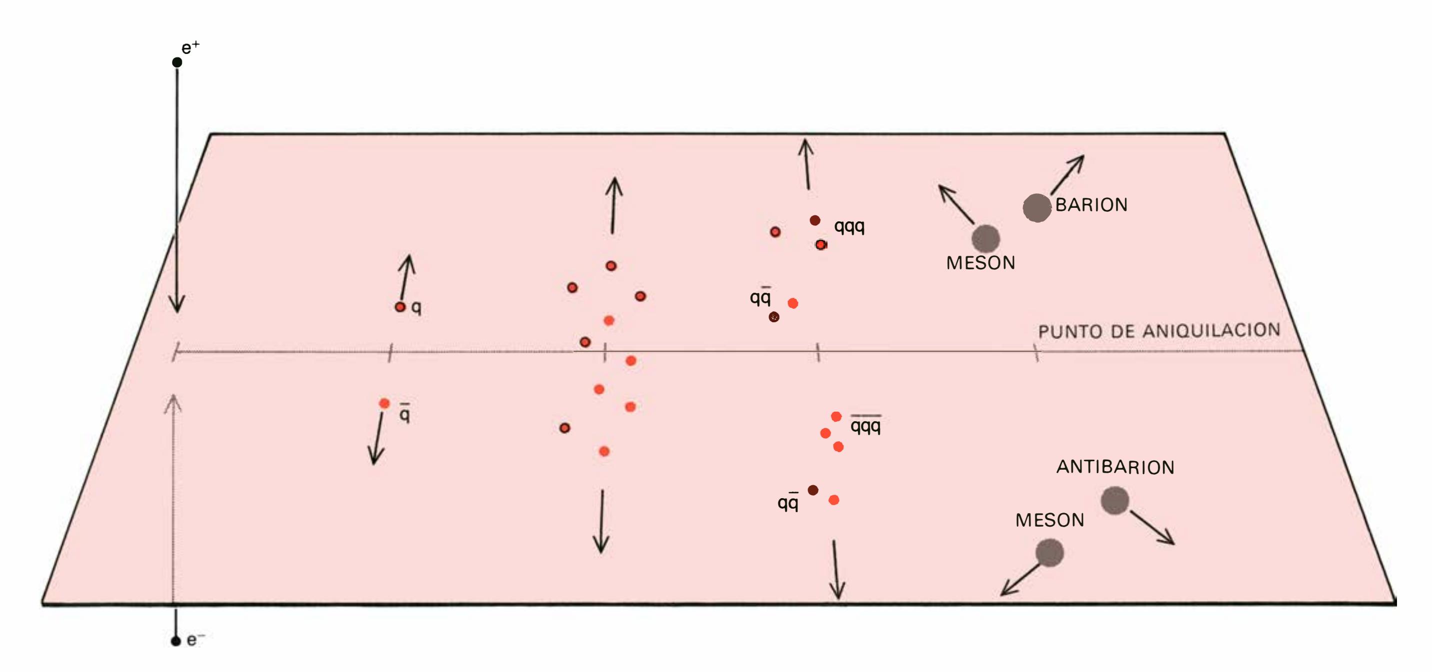
La energía de los haces almacenados puede ajustarse entre 1,2 GeV y 4 GeV. Una de las principales razones que inducen a emplear anillos de almacenamiento es que toda esa energía se halla disponible para la creación de nuevas partículas. Cuando un electrón y un positrón chocan, ambas partículas se aniquilan y toda su masa y energía cinética se convierte en un estado de energía pura. Así, la máquina es capaz de crear conjuntos de partículas con masa total de hasta 8 GeV.

La energía liberada en la aniquilación queda confinada en una pequeña región del espacio, de dimensiones de 10^{-15} a 10^{-14} centímetros. En un breve intervalo inconmensurable –el tiempo necesario para que la luz atraviese dicha región–, la energía se transforma en partículas materiales, que entonces se alejan del punto de aniquilación. Hay pocas restricciones sobre las clases de partículas que pueden crearse a partir de este denso paquete de energía. Una exigencia obvia es que la energía debe conservarse: las masas en reposo de las partículas creadas no pueden ser mayores que la suma de las energías del positrón y el electrón. Los números cuánticos también deben conservarse, pero esta condición es menos restrictiva que lo que pueda parecer en un principio.

La aniquilación de un electrón y un positrón se produce mediante la fuerza electromagnética; el producto inmediato es un fotón, un cuanto de energía electromagnética. Este fotón se desintegra tan rápidamente que nunca puede detectarse, ni tan sólo en principio (se llama una partícula virtual); sin embargo, determina las propiedades de todos los subsiguientes estados del sistema. Si todos los productos de desintegración del fotón virtual se detectaran, y se sumaran sus números cuánticos, los totales deberían ser iguales a los números cuánticos del fotón. La partícula psi puede crearse en estos sucesos porque sus números cuánticos son precisamente los del fotón: tiene una unidad de momento angular, pero todos sus números cuánticos aditivos –los determinables por simple suma de los números cuánticos de los quarks constituyentes– son cero. La razón de que ello sea así se hace evidente cuando se considera la composición en quarks de la psi. Esta partícula consta de un quark ligado a su propio antiquark; y todos los números cuánticos, excepto el momento angular, se cancelan exactamente.

En este contexto, el hecho de que se conserven las propiedades del fotón, que había parecido una condición restrictiva, confiere una libertad casi com-

LAS CAMARAS DE CHISPAS en el detector magnético consisten en capas concéntricas de finos hilos de aluminio; cuando se dispara el detector, se aplica una tensión elevada entre las capas. Aquí se muestran sólo dos capas; en realidad, cada cámara tiene cuatro capas. Hay cuatro cámaras, con un total de unos 100.000 alambres. Cuando la cámara de chispas está en marcha, los alambres se hallan inmersos en un gas noble que se ioniza con el paso de una partícula cargada eléctricamente. Los átomos ionizados que deja proporcionan una trayectoria conductora para una chispa entre los alambres de las capas de carga opuesta, marcando, por tanto, la posición en la que la partícula atravesó la cámara cilíndrica. La información posicional se registra mediante un sistema de "varas magnetostrictivas" colocadas perpendicularmente a los alambres de aluminio en el extremo de la cámara. El impulso de corriente que fluye a través del alambre activo, cuando se dispara una chispa, da lugar a un pulso en la vara magnetostrictiva, pudiendo detectarse y medirse su tiempo.



LA ANIQUILACION de un electrón y un positrón da lugar a igual número de partículas y antipartículas. Cuando se crean hadrones como resultado de una colisión, el producto inicial es un quark y un antiquark de la misma clase. A partir de la energía proporcionada a esta pareja, se pueden materializar otros quarks y antiquarks, que se agrupan en pares y tríos para formar hadrones. Sólo pueden detectarse los hadrones.

10 KILOMETROS
10 METROS
1 CENTIMETRO
10^{-3} CENTIMETRO
10^{-6} CENTIMETRO
10^{-9} CENTIMETRO
10^{-12} CENTIMETRO
10^{-15} CENTIMETRO
ZONA DE ANIQUILACION
DISTANCIA DESDE EL PUNTO DE ANIQUILACION
ESTABLE
ESTABLE
ESTABLE
ESTABLE
ZONA DE DETECTORES
DESINTEGRACIONES DEBILES
DESINTEGRACIONES ELECTROMAGNETICAS
DESINTEGRACIONES FUERTES
DESINTEGRACION MUONES
DESINTEGRACION NEUTRONES
10^{-24} 10^{-21} 10^{-18} 10^{-15} 10^{-12} 10^{-9} 10^{-6} 10^{-3} 1 10^{3}
DESINTEGRACIONES DE LAS PSI
VIDA MEDIA DE LA PARTICULA (SEGUNDOS)

LAS DESINTEGRACIONES EN CASCADA que siguen a una aniquilación electrón-positrón convierten la gran variedad de partículas creadas en unas cuantas especies estables o casi estables. La mayoría de hadrones que pueden desintegrarse mediante las interacciones fuertes han completado sus desintegraciones en 10^{-23} segundos, cuando han recorrido sólo unos 10^{-13} centímetros desde el punto de la colisión. La partícula psi es una excepción: se desintegra mediante la interacción fuerte, si bien tarda unos 10^{-20} segundos. Los productos de las desintegraciones fuertes son hadrones de más larga vida, algunos de los cuales se desintegran a continuación electromagnéticamente, dando otros hadrones y rayos gamma (γ). Las desintegraciones débiles son las más lentas, pero la mayoría de ellas se han completado antes de que las partículas hayan recorrido más de un centímetro. Después de las desintegraciones débiles, sólo sobreviven leptones, rayos gamma y los hadrones de mayor vida. Los detectores se pueden arrimar hasta escasos centímetros de la zona de aniquilación, lo que comporta que se detecten los productos de la colisión después de que la mayoría de desintegraciones se hayan acabado. Los rayos gamma indican que una partícula se ha desintegrado electromagnéticamente, y los leptones son signos de una desintegración débil. Los mesones *K* reflejan la presencia de un hadrón extraño.

pleta para la creación de cualquier partícula, siempre que esté acompañada por su propia antipartícula. En la psi, el par partícula-antipartícula es interno, pero no necesitaría serlo. Por ejemplo, podría producirse un mesón encantado en combinación con otro mesón de número cuántico de encanto opuesto. Dada la energía de haz suficiente, puede formarse casi cualquier estado de la materia (y antimateria). Las colisiones del SPEAR liberan energía capaz de producir hasta 50 piones (mitad π^+ y mitad π^-), o cuatro protones junto con cuatro antiprotones o incluso un átomo completo de helio y un átomo de antihelio.

Como las partículas salen de la región de aniquilación a una velocidad próxima a la velocidad de la luz, suelen desintegrarse a través de varias generaciones en partículas hijas más ligeras y de más larga vida. La mayoría de los hadrones pesados han sufrido desintegraciones fuertes a lo largo del tiempo en que han viajado 10^{-13} centímetros, distancia aproximadamente igual al diámetro del hadrón. Los productos de estas desintegraciones son hadrones más ligeros, tales como piones, mesones *K*, protones y neutrones. Los hadrones que se desintegran electromagnéticamente en general no llegan a los 10^{-6} centímetros de su viaje antes de emitir un fotón y convertirse en hadrones de masa menor. La mayoría de partículas que se desintegran débilmente, como el barión lambda extraño, se desplazan unos pocos centímetros antes de transformarse en los hadrones de masa menor, como protones, neutrones y piones, junto con leptones. A diez centímetros del punto de la interacción, la mayoría de las partículas supervivientes son protones y neutrones, piones y mesones *K* cargados eléctricamente, rayos gamma y leptones. En unos pocos metros, los piones y los mesones *K* se desintegran en leptones, dejando sólo protones, neutrones, rayos gamma, electrones, muones y neutrinos. Cuando se ha recorrido un trayecto mucho mayor, se desintegran los neutrones y los muones.

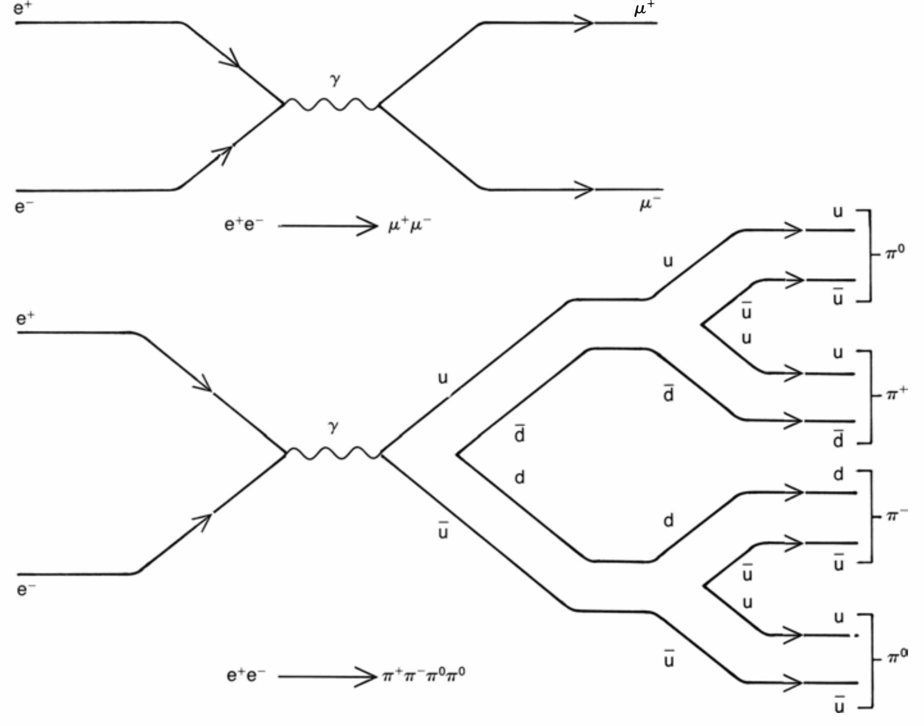
La meta del experimentador es detectar tantas partículas de éstas cuantas pueda y extraer de ellas la máxima cantidad de información. En la práctica, el detector sólo puede arrimarse hasta unos centímetros de la región de interacción; y así no hay esperanza de ver partículas antes de que las desintegraciones fuertes y electromagnéticas hayan sucedido. En el otro extremo de la escala, los detectores de más de unos pocos metros de diámetro resultan demasiado caros y engorrosos para ser prácticos. Incluso dentro de este rango limitado de distancias sólo pueden detectarse generalmente las partículas cargadas y, en algunos casos, los rayos gamma. No obstante, con frecuencia se puede reconstruir, a partir de las partículas detectadas, la cadena de sucesos que dieron lugar a ellas. Por ejemplo, la presencia de mesones *K* entre los productos de la desintegración puede interpretarse como un signo de que se crearon partículas extrañas; los rayos gamma indican que ha habido una desintegración electromagnética y los leptones dan cuenta del paso de una partícula que se desintegra débilmente.

El detector en el que se descubrió la psi rodea una de las regiones de interacción del SPEAR. Fue realizado por físicos del SLAC y del Lawrence Berkeley Laboratory, mientras el SPEAR estaba en fase de construcción. Algunos miembros y pequeños grupos de ambos laboratorios se encargaron de los componentes del detector y de sus programas de ordenador.

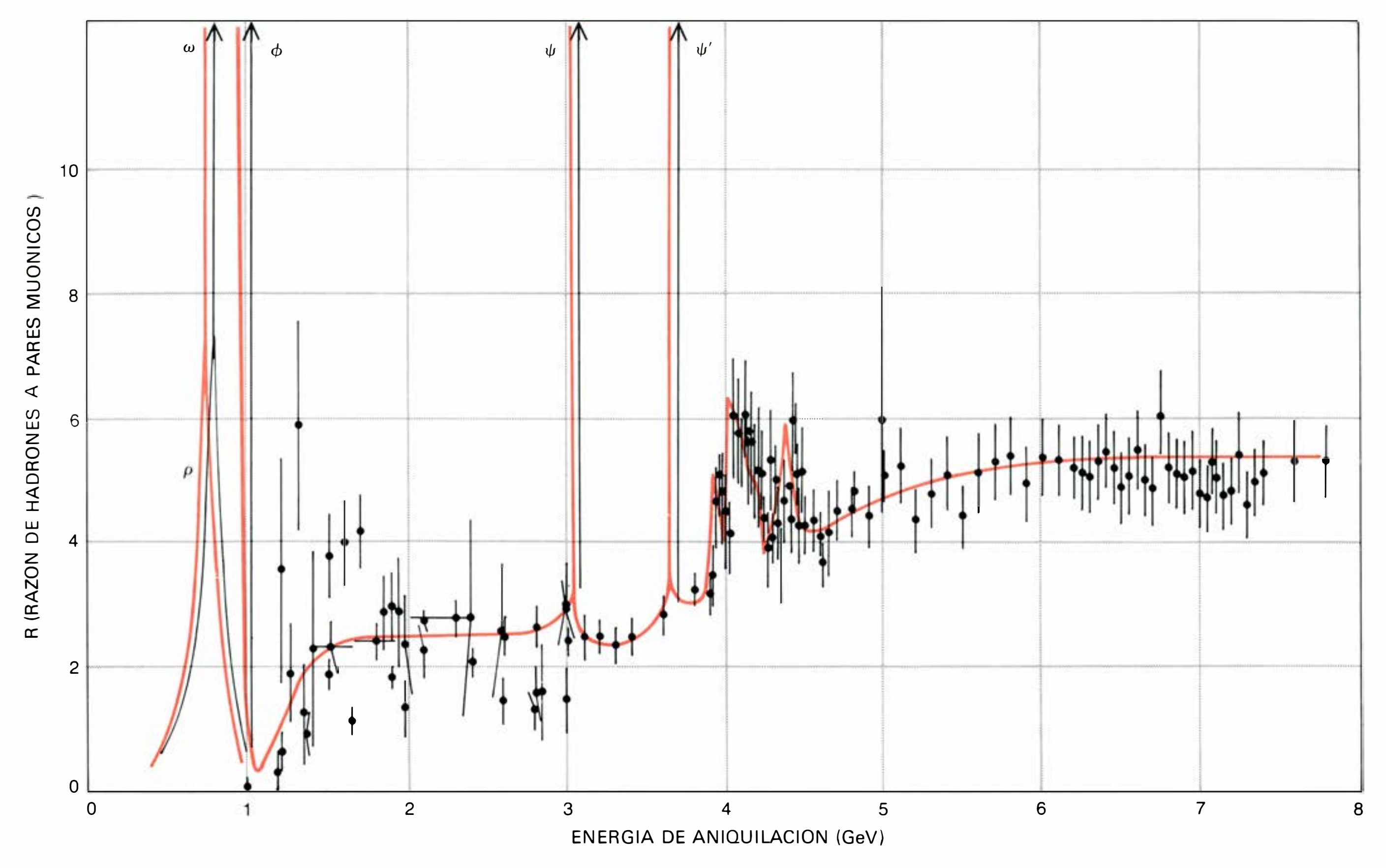
La base del detector es un imán solenoidal de 150 toneladas, cuyo volumen de campo útil es un cilindro de tres metros de diámetro por tres metros de largo. Dentro del campo, hay cuatro capas cilíndricas de cámaras de chispas, que registran las trayectorias de las partículas cargadas cuando pasan a través del detector. La cámara de chispas contiene un total de 100.000 alambres de aluminio, sensibles a la ionización momentánea de un gas inerte causada por el paso de una partícula. El campo magnético curva las partículas cargadas; midiendo la curvatura de sus trayectorias, se puede determinar el momento de las partículas.

Más allá de las cámaras de chispas hay contadores de centelleo, en donde se detectan fotoeléctricamente los breves destellos de luz emitidos cuando una partícula penetra en una faja de plástico fosforescente. El instante de los centelleos puede medirse con precisión, y en consecuencia puede calcularse el tiempo de vuelo y la velocidad de la partícula. Conociendo la velocidad de la partícula y su momento, se puede deducir su masa en reposo, que es un dato importante para su identificación. La comparación de velocidad y momento es el mejor medio que tenemos para distinguir entre los tres hadrones cargados que sobreviven lo suficiente como para ser detectados, a saber: piones, mesones *K* y protones.

Un último sistema periférico de contadores de centelleo y cámaras de chispas está protegido del interior del detector por capas filtrantes de plomo y hierro. Las diferentes capacidades de las distin-



LOS SUCESOS DE ANIQUILACION pueden dividirse en dos grandes categorías: los que dan lugar a un par de leptones, verbigracia, un muón positivo y negativo, y los que conducen a la creación de un quark y un antiquark, que pueden aparecer como un conjunto de hadrones. Si leptones y quarks tienen una distribución puntual de carga eléctrica, entonces pueden calcularse los ritmos de producción; los ritmos se determinan por el número de clases de leptones y quarks y por los cuadrados de sus cargas. Las mediciones del ritmo de producción de los hadrones pueden someter a prueba por tanto los modelos de quarks, discriminando, por ejemplo, entre los que incluyen quarks encantados y los que no. En la práctica no se miden ritmos absolutos de producción; sí se determina la razón de la producción de hadrones a la producción de muones, designada por la letra *R*.



LA RAZON *R* aumenta desde unos 2,5 a energías por debajo de 3000 millones de electrón-voltios (3 GeV) hasta unos 5,5, a energías por encima de 5 GeV. La interpretación más plausible del aumento es la que sostiene que en este intervalo se ha cruzado el umbral de creación de una nueva clase de quarks. Más que el aumento total de la razón, lo que debe destacarse es que la curva tiene una serie de picos altos y extraordinariamente estrechos. Los tres picos agrupados cerca de 1 GeV representan la producción de los mesones rho (ρ), omega (ω) y phi (φ), cuyas combinaciones de quarks componentes son: $u\bar{u}$, $d\bar{d}$ y $s\bar{s}$. El pico a 3,1 GeV es el mesón psi (ψ) formado por un par quark-antiquark $c\bar{c}$. La psi se descubrió en 1974 mientras se obtenía una medición de *R*; la psi prima (ψ′), un estado excitado del sistema $c\bar{c}$, se halló 10 días después. El pico más ancho a 4 GeV representa varias partículas de vida corta; se considera el umbral de creación de un par de mesones encantados.

tas partículas para atravesar estos metales nos permite discernir entre hadrones y leptones y, dentro de los leptones, entre electrones y muones.

A lo largo de la mayor parte del tiempo, los haces de electrones y positrones se cruzan entre sí sin interaccionar. A pesar de que cada paquete contiene 100.000 millones de partículas, los leptones son tan pequeños –su tamaño de momento parece ser realmente inconmensurable– que la probabilidad de una colisión es baja. Por otro lado, la enorme celeridad a que se mueven los haces, que recorren el anillo más de un millón de veces por segundo, permite lograr ritmos adecuados de recolección de datos.

El detector se dispara cuando dos o más partículas cargadas, provenientes de un entorno de la región de interacción, alcanzan los contadores de centelleo simultáneamente al paso de los haces a través del detector. Cuando se registra un tal suceso, toda la información de las cámaras de chispas y de los contadores de centelleo se manda a un ordenador, que la registra en cinta magnética. El detector se dispara cerca de una vez por segundo, pero sólo unos pocos por ciento de estos sucesos provienen de aniquilaciones electrón-positrón. El resto están causados por rayos cósmicos y por interacciones entre haces de partículas y moléculas del gas residual de la cámara de vacío.

Los datos recogidos en la cinta se analizan con la ayuda de otro ordenador. Los sucesos contaminantes se identifican y rechazan; para los sucesos más interesantes, se reconstruyen las trayectorias de las partículas. En un suceso típico en que se hallen implicados hadrones pueden detectarse cuatro o cinco partículas; se miden sus momentos y velocidades; y, si ello es posible, se identifican las partículas.

En un instante dado, hay unos 35 físicos trabajando en el funcionamiento del detector y en el análisis de los datos. Aunque el número varía a medida que estudiantes e investigadores ocasionales vienen y van, sin la dedicación y el talento de muchos individuos que componen el equipo este gran aparato de detección y análisis no podría operar.

Cuando se descubrió la partícula psi, en 1974, no estábamos buscando partículas encantadas sino que nos hallábamos inmersos en una investigación mucho más amplia sobre la estructura de los hadrones. Pretendíamos comprobar la hipótesis de que los hadrones están formados por partículas puntuales, cargadas eléctricamente, tales como se supone que son los quarks.

La electrodinámica cuántica, una teoría que ha sido verificada con gran precisión, puede predecir los ritmos de creación de partículas en las aniquilaciones electrón-positrón, siempre que las partículas tengan una distribución puntual de carga eléctrica. El ritmo es proporcional a la suma de los cuadrados de las cargas de las partículas. Esta hipótesis se comprobó para los electrones y los muones, con perfecto acuerdo entre experimentación y teoría. Se supone que los quarks se comportan exactamente igual que los leptones en las interacciones electromagnéticas, y así la misma regla debería predecir también su ritmo de producción.

No podemos observar directamente los quarks creados a consecuencia de la colisión electrón-positrón, pero ello no impide medir su ritmo de producción. Cuando los hadrones aparecen entre los productos finales de una aniqui-

lación, se supone que todos provienen, en definitiva, de un único par quark-antiquark. Por tanto, el ritmo de producción de quarks se obtiene simplemente con el número de sucesos en los que aparece algún hadrón. En la práctica no se mide el ritmo absoluto. Es más conveniente medir la razón de la producción de hadrones a la producción de muones, una razón que convencionalmente se designa por *R*. Si la hipótesis de los quarks es válida, *R* debe ser igual a la suma de los cuadrados de las cargas de los quarks, y debe ser aproximadamente constante e independiente de la energía de colisión.

La medición experimental de *R* sigue un proceso directo. A las energías del anillo de almacenamiento del SPEAR, los sucesos en que aparecen hadrones tienen casi siempre tres o más partículas en el estado final, mientras que los muones se producen estrictamente a pares que abandonan la región de interacción con momentos iguales y opuestos. Así pues, puede obtenerse una excelente estimación de *R* sumando simplemente el número de sucesos en que se detectan más de dos partículas y dividiéndolo por el número de sucesos de dos cuerpos que salen en direcciones opuestas.

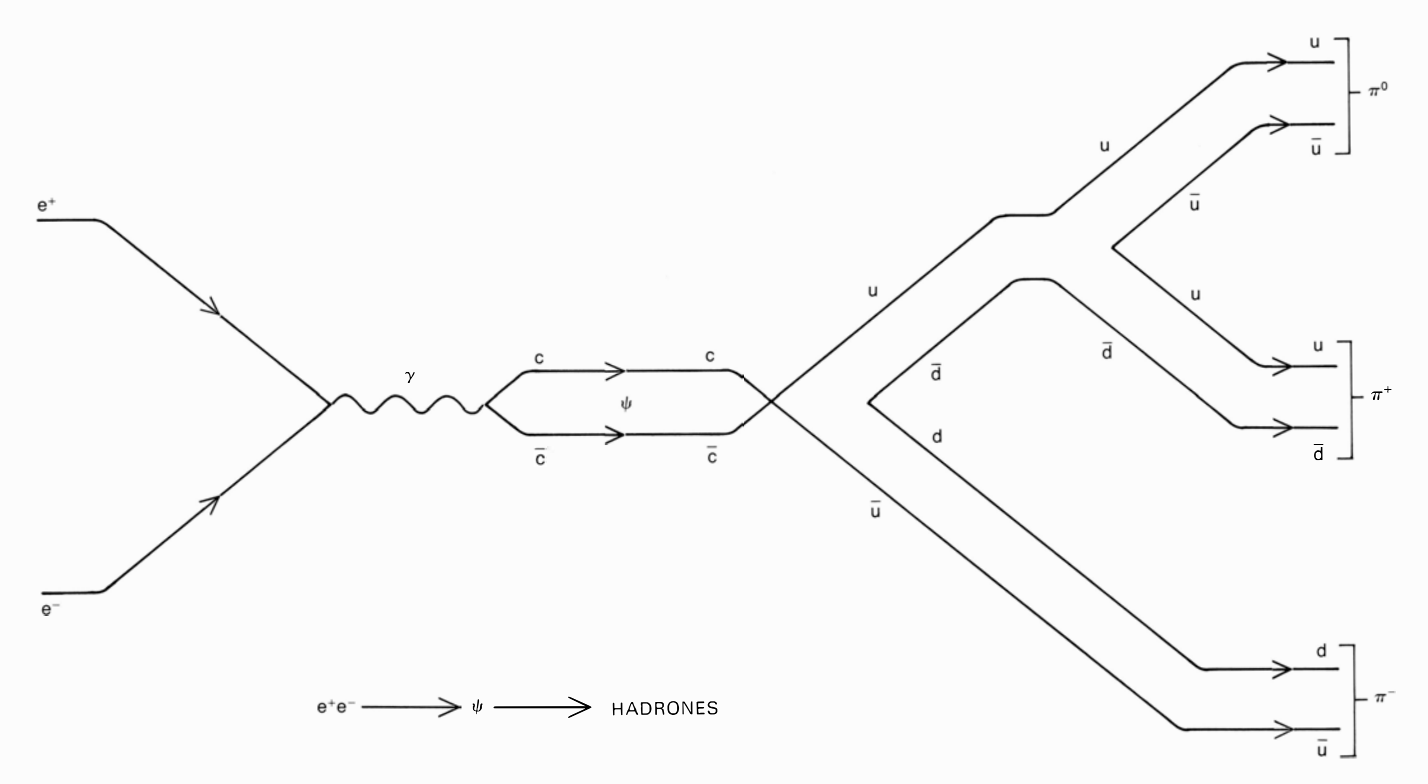
La razón *R* es una prueba importante de las teorías de los quarks. Debido a que *R* se determina por la suma de los cuadrados de las cargas de los quarks, puede discriminar entre variantes del modelo de quarks que tengan distinto número de quarks o distintas asignaciones de cargas.

La determinación experimental de los valores de *R* tiene una historia interesante; éstos se han mostrado coherentemente demasiado grandes para las teorías de moda de cada momento. En el modelo original de Gell-Mann y Zweig con tres quarks, *R* debía haber sido igual a $(2/3)^2+(1/3)^2+(1/3)^2$ que da 2/3. Este valor de *R* implica que, para cada dos sucesos que contengan hadrones, deben haberse producido tres pares de muones. Las primeras mediciones de *R* se hicieron con el anillo de almacenamiento llamado ADONE, de Frascati, Italia. Estas mediciones, realizadas todas a energías por debajo de 3 GeV, dieron valores, no de 2/3, sino de aproximadamente 2; en otras palabras, había dos veces más sucesos hadrónicos que pares de muones.

Poco después, la teoría alcanzaba al experimento. Se propuso, y hoy es ampliamente aceptado, que cada uno de los quarks debe presentarse en tres variedades a fin de dar cabida a un nuevo número cuántico, generalmente llamado color. Mientras no pueda aislarse un quark, el número cuántico de color seguirá permaneciendo oculto. Se ha comprobado que la idea del color es excepcionalmente fructífera para explicar las interacciones de los quarks; se trata, empero, de un número cuántico puramente "interno" y no predice hadrones adicionales. En realidad, una de las pocas maneras que tiene de influir en los resultados experimentales es en el valor de *R*. Triplicando el número de los quarks, *R* también se triplica, es decir, llega a un valor de 2.

Apenas si hubo tiempo para celebrar ese resultado. A principios de la década de los 70 se hicieron otras dos mediciones de *R*, a energía más alta, con el anillo de almacenamiento del Cambridge Electron Accelerator. Mostraron que, a 4 GeV, *R* vale 4 y, a 5 GeV, sube hasta casi 6.

El Cambridge Electron Accelerator se cerró poco después de la realización de estas mediciones, pero el programa inicial de experimentos en el SPEAR, que empezó a finales de 1973, confirmó y extendió los resultados. Se midió *R* a varias energías, separadas por intervalos de anchura adecuada. Por debajo de unos 3 GeV, el valor es 2,5, en razonable acuerdo con el valor 2 esperado si hay tres quarks y cada uno aparece en tres colores. Por encima de 4 GeV, sin embargo, la razón sube rápidamente, y, rebasando los 5 GeV, alcanza un valor próximo a 5,5. No cabe duda de que algún nuevo fenómeno debe hacer su aparición. Tal vez ni siquiera el encanto baste para explicar el comportamiento



LA PARTICULA PSI puede crearse en una aniquilación electrón-positrón porque consta de un quark y un antiquark del mismo tipo. El producto inmediato de una colisión electrón-positrón es siempre un fotón, pero éste se desintegra invariablemente en otras partículas antes de que pueda observarse. El fotón tiene una unidad de momento angular de espín, pero todos sus otros números cuánticos son nulos; cualquier partícula obtenida del fotón debe tener los mismos números cuánticos. La psi es una tal partícula; los mesones rho, omega y phi son otras. La psi se desintegra por la aniquilación del quark y antiquark encantados, dando lugar a hadrones ordinarios como los tres piones mostrados aquí.

de la curva. Añadiendo el quark encantado al modelo (con tres colores), se sube el valor esperado de R sólo hasta 3,3.

Estos resultados desconcertantes despertaron gran interés, pero quedaron pronto ensombrecidos ante otro descubrimiento registrado en la curva R: el pico que señalaba la presencia de la partícula psi. En una gráfica de R, la psi aparece como un pico estrecho de enorme altura a una energía de 3,095 GeV. Si la estructura del pico pudiera clarificarse perfectamente, se vería que se levanta sobre el nivel del fondo en un factor de 3000; en otras palabras, a 3,095 GeV es 3000 veces más probable que se creen hadrones que a las energías próximas. Se denomina resonancia a ese aumento en la producción de hadrones. Cuando la energía de colisión del electrón y el positrón se ajusta exactamente a la masa de la partícula psi, la probabilidad de que se cree una psi es elevada. La desintegración subsiguiente de la psi produce los hadrones que se detectan como una resonancia. A energías sólo ligeramente más altas o más bajas que la de la resonancia, la energía y el momento ya no pueden ajustarse para producir sólo la psi, y el ritmo de producción de hadrones baja.

Más destacable incluso que la altura de la resonancia psi es su extremada estrechez, propiedad que está relacionada con la vida media de la partícula. De hecho, sólo a partir de las mediciones de la anchura de la resonancia puede deducirse la vida media. Parte de la anchura observada proviene de la resolución finita del aparato experimental. Sin embargo, ni siquiera con un aparato perfecto podría el pico quedar asociado a una energía únicamente definida. El rango de energías sobre las que se distribuye el pico puede interpretarse en mecánica cuántica como representante de la incertidumbre residual en la energía real de la partícula, incertidumbre que se relaciona con la vida media. A menor vida media, mayor incertidumbre. Es como si no hubiese tiempo suficiente de medir con precisión su energía antes de que la partícula se desintegrara. La mayoría de resonancias en esta región de masas son picos muy anchos, de cientos de millones de electrón-voltios de ancho: la anchura de la resonancia psi es de 67.000 electrón-voltios, es decir como un 0,002 por ciento de su energía total.

El descubrimiento inesperado de ese pico estrecho en la gráfica de la razón R indujo un cambio en la rutina de funcionamiento de los anillos de almacenamiento. En vez de estudiar la producción de hadrones a unas pocas energías muy separadas, se hicieron exploraciones de alta resolución con el incremento mínimo posible entre las energías estudiadas. Con gran diferencia, el mejor estudio de alta resolución realizado en el SPEAR fue el primero: 10 días después del descubrimiento de la psi se encontró una segunda partícula de masa 3,684 GeV, estrechamente relacionada con la psi. A esta segunda partícula se la llamó psi prima (ψ'). Al igual que la psi, la psi prima es neutra eléctricamente y tiene una unidad de espín; no tiene vida media tan larga como la psi, pero la anchura de la resonancia, de 228.000 electrón-voltios, es aun mucho más estrecha que la mayoría.

Ulteriores exploraciones de la curva R revelaron un tercer hecho en la vecindad de 4,1 GeV. Este pico es mucho menos claro; es un abultamiento con una anchura de unos 200 MeV. En un principio no quedó claro cómo había que interpretarlo.

Un estado ligado de un quark encantado y un antiquark encantado tiene mucho en común con los átomos simples, por ejemplo, el átomo de hidrógeno. Inmediatamente antes del descubrimiento de la psi, Thomas Appelquist y H. David Politzer, de Harvard, propusieron el nombre de "charmonium" para dicho estado ligado. El nombre implica una analogía con otro sistema exótico parecido al átomo, el positronium, que es un estado ligado de un electrón y un positrón.

De acuerdo con el modelo del charmonium, el quark y el antiquark de la partícula psi tienen sus espines paralelos, pero carecen de momento angular orbital. La psi prima es un estado excitado de este sistema, análogo a uno de los estados excitados del átomo de hidrógeno. En este estado, los espines de los quarks son también paralelos y no hay momento angular orbital. Los espines paralelos de los quarks se combinan para dar una unidad de espín a cada mesón. Esta disposición de los espines se conoce en física atómica como una configuración "orto", y las partículas psi y psi prima pueden denominarse ortocharmonium.

Siguiendo el razonamiento de los principios de la física atómica, Appelquist y Politzer pudieron predecir formas adicionales del charmonium. Por ejemplo, debía haber estados sin momento angular orbital y con los espines de los quarks antiparalelos. Los espines en direcciones opuestas se restan, con el resultado de que el conjunto de la partícula tiene momento angular cero. Esta configuración se denota con el prefijo "para" y tiene una energía total más baja que la configuración orto. Así debe haber un estado paracharmonium con masa ligeramente menor que la de la psi y otro con masa ligeramente menor que la de la psi prima. El quark y el antiquark deben también combinarse con espines paralelos y con una unidad de momento angular orbital. Se esperan tres de tales estados, conocidos en conjunto como charmonium de onda p (otro término tomado de la física atómica). Sus masas deben estar todas entre la masa de la psi y la masa de la psi prima.

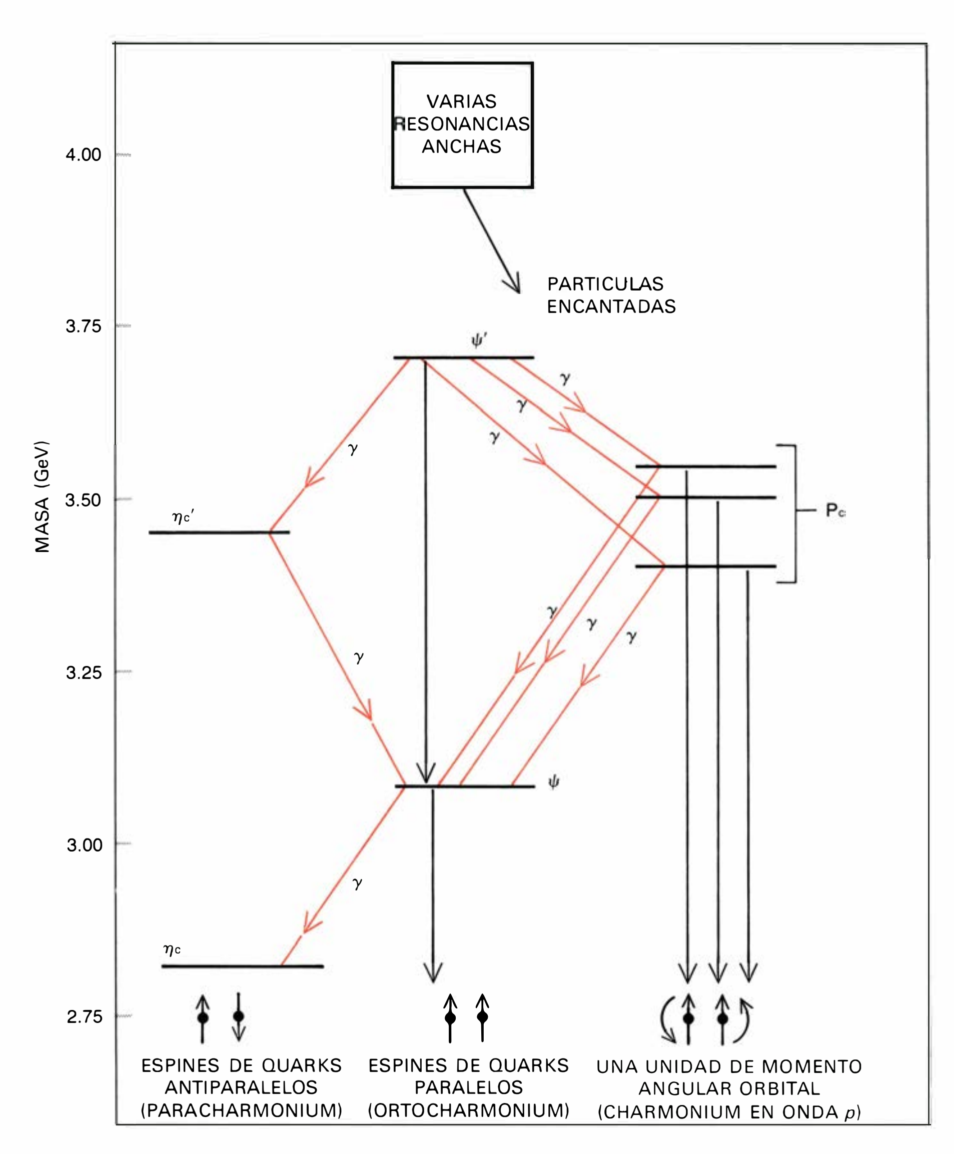
Como los números cuánticos del fotón deben conservarse, las aniquilaciones electrón-positrón sólo pueden dar lugar a pares quark-antiquark en la configuración orto; de este modo, ni el paracharmonium ni el charmonium de onda p pueden observarse como picos en la razón R. Sin embargo, debe ser posible producirlos mediante transiciones electromagnéticas de los estados psi y psi prima. Una vez más, estas transiciones son análogas a ciertos sucesos de la física atómica, en los cuales un átomo puede pasar de un estado excitado a otro estado de energía menor por emisión de un fotón. En las transiciones atómicas los fotones emitidos están típicamente en la región visible del espectro electromagnético; en las transiciones entre los estados del charmonium, los fotones son rayos gamma de alta energía.

El enfoque más directo para hallar tales estados estriba en producir abundantemente la partícula psi prima, y buscar entre sus productos de desintegración los rayos gamma mencionados. Tales experimentos se pusieron en marcha en el SPEAR y en el anillo de almacenamiento llamado DORIS, asociado al German Electron Synchrotron, instalado cerca de Hamburgo.

A la hora de reconocer los rayos gamma, la principal dificultad con que topamos radica en distinguirlos de otros muchos fotones emitidos en procesos no relacionados. Una pista que arroja luz en esta discriminación es que todos los fotones emitidos durante las transiciones entre dos estados dados tienen la misma energía, de manera que deben aparecer por encima de los sucesos contaminantes o de fondo, los cuales están dotados de una amplia gama de energías. En los primeros experimentos el fondo se redujo aun más, mediante la selección, para su examen, de sólo aquellos acontecimientos que parecían ajustarse a una determinada sucesión de desintegraciones. Decidimos estudiar los sucesos en los que una partícula psi prima emite un fotón y se desintegra a un estado intermedio, el cual emite luego

otro fotón y da una partícula psi. Pudimos así excluir de nuestras consideraciones todos los sucesos, excepto aquellos en los que podía deducirse la presencia de una psi y en los que había exactamente dos rayos gamma. Tales metodos de "cortar" los datos son corrientes en física de alta energía; permiten reducir, a un puñado de sucesos interesantes, cientos de miles de interacciones registradas.

Los primeros resultados fueron presentados por un grupo de experimentadores que trabajaban en el DORIS; encontraron 14 sucesos que podían ajustarse a ese criterio. Poco después, nuestro grupo del SPEAR aportó algunos otros candidatos. Estos pocos sucesos contenían una importante información que no era una pura consecuencia del proceso de selección: las energías de los dos fotones. En la mayoría de los sucesos, uno de los rayos gamma tenía una energía de unos 170 MeV; restando esta energía a la masa de la psi prima, dedujimos que la psi prima se desintegraba, al menos algunas veces, a un estado intermedio de masa cercana a 3,5 GeV. A continuación, el estado intermedio se desintegraba en la partícula psi. La partícula de 3,5 GeV resultó ser uno de los estados de charmonium en onda *p*, que están dotados de una unidad de momento angular orbital. El grupo de DORIS le llamó P_c. A partir de entonces se han obtenido en el SPEAR y en el DORIS más ejemplos de esta "desintegración en cascada" y se han encontrado los dos estados de onda *p* restantes, cuyas masas son ligeramente distintas.



EL ESPECTRO DE PARTICULAS que incluye la psi consta de siete estados, por lo menos. Todos ellos están formados por un quark encantado y un antiquark encantado, pero se distinguen por varias combinaciones del espín y del momento angular orbital. Sólo la psi y la psi prima pueden crearse directamente a partir de aniquilaciones electrón-positrón, porque únicamente estos estados tienen los mismos números cuánticos que el fotón. Los espines de los quarks en la psi están orientados en la misma dirección y se suman para dar al mesón una unidad de espín. Se forman dos estados de energía algo menor, llamados η_c y η'_c, cuando los espines de los quarks son antiparalelos. Se crean tres estados de masa intermedia, llamados excitaciones de onda *p*, cuando un par de quarks con espines paralelos adquieren una unidad de momento angular orbital. El estado ligado de un quark encantado y un antiquark encantado recuerda a un átomo simple; a todos los estados de este sistema se les ha dado el nombre de "charmonium". Como en el espectro de niveles de energía de un átomo, un estado puede transformarse con otro de menor energía por emisión de un fotón.

Nuestro grupo del SPEAR acometió una segunda gran investigación de transiciones de rayos gamma. Buscábamos pruebas de desintegraciones de la psi prima en las que el estado final contuviera un conjunto determinado de hadrones y un solo rayo gamma. Estos productos de desintegración indicarían un estado intermedio del charmonium, que no se desintegrara en una partícula psi, sino en hadrones ordinarios. Fuimos recompensados con el descubrimiento de que, de nuevo, los fotones se agrupaban en ciertas energías discretas, dos de las cuales coincidían con las energías observadas en las desintegraciones en cascada.

El año del descubrimiento de la partícula psi, había aparecido ya un espectro de estados estrechamente relacionados. Estos incluían no sólo las excitaciones de onda *p*, sino también el estado de paracharmonium situado justo debajo de la psi prima. Dos grupos de experimentadores del DORIS llegaron a comprobar el estado de paracharmonium debajo de la propia psi; éste es el estado de charmonium de menor masa. Por último, en 1976, la mayoría de picos del espectro de rayos gamma se observaron en una misma medición, que realizaron físicos pertenecientes a cinco instituciones, a saber: Universidad de Maryland, Universidad de California en San Diego, Universidad de Pavia, Universidad de Princeton y Universidad de Stanford. Este espectro de rayos gamma es como el espectro de líneas de un átomo ordinario, con una diferencia espectacular: la energía de un fotón típico en una transición atómica suele ser de unos pocos electrón-voltios, mientras que los rayos gamma emitidos en las transiciones del charmonium tienen energías de varios centenares de millones de electrón-voltios.

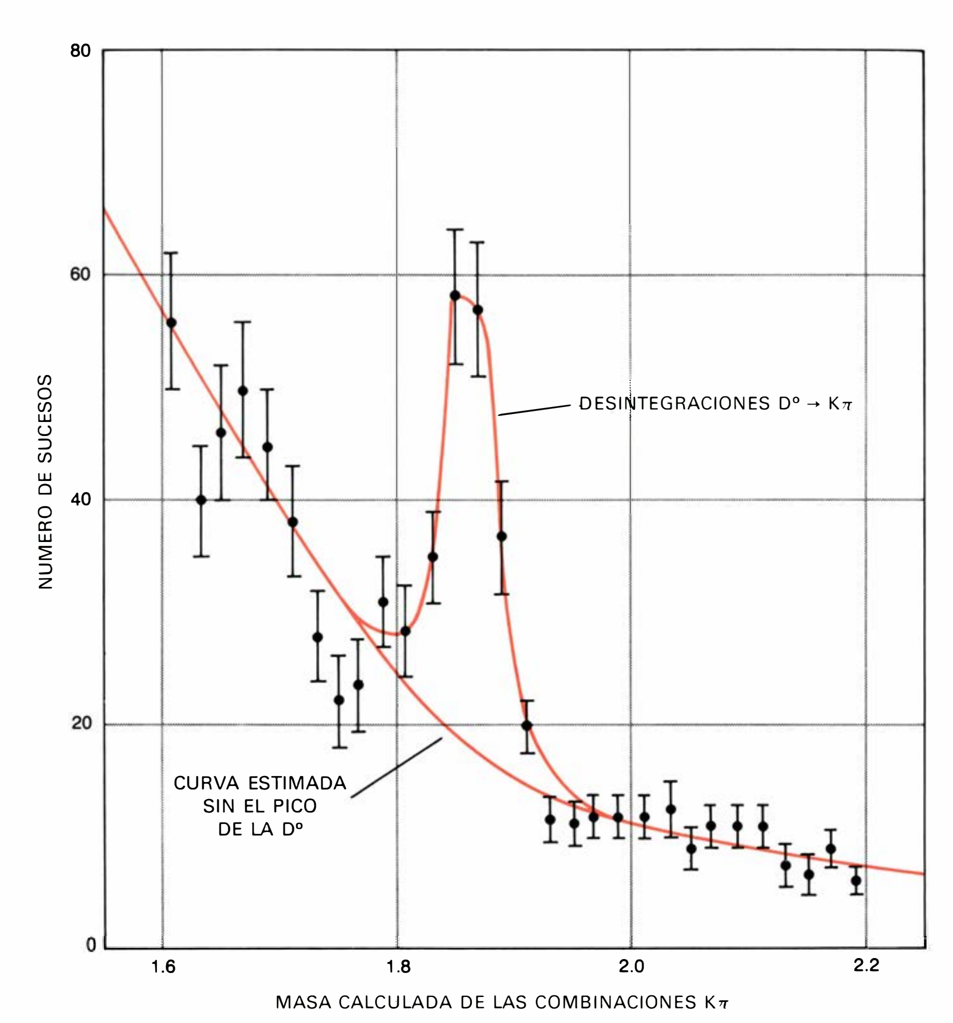
El espectro de estados del charmonium representa una prueba vigorosa de que la psi y sus congéneres constan de una partícula ligada a su propia antipartícula; pero no hay certeza de que el número cuántico que detenten esas partículas sea el encanto. Esta propiedad permanece oculta en todos los estados del charmonium. Se manifiesta solamente en los hadrones que incluyen un quark o un antiquark encantados combinados con quarks sin encanto.

En las colisiones electrón-positrón, tales hadrones encantados sólo pueden crearse a pares. Por tanto, la mínima energía de aniquilación a la que puede aparecer el encanto equivale a dos veces la masa de la partícula encantada más ligera. Puede estimarse esta energía umbral a partir de un examen de la razón *R*. Una cota inferior es la masa de la psi prima: si el umbral estuviese por debajo de esta masa, entonces la propia psi prima se desintegraría rápidamente dando un par de partículas encantadas, y la resonancia sería mucho más ancha de lo que es. Un aumento significativo de *R* con varios picos anchos, justo por encima de la masa de la psi prima, sugiere la creación de estados de charmonium altamente excitados, que se desintegran fuertemente en partículas con encanto no nulo. Así el umbral parece estar entre 3,7 GeV y unos 4 GeV, lo cual implica que la partícula encantada más ligera tiene una masa entre 1,85 GeV y unos 2 GeV.

Al igual que las partículas pesadas extrañas, las encantadas pesadas pueden desintegrarse fuertemente conservando el encanto mediante el tránsito del quark *c* o el antiquark $\bar{c}$ a un estado encantado más ligero. Las partículas encantadas más ligeras, sin embargo, se desintegrarán sólo mediante interacciones débiles, y, en comparación, tendrán una vida media más larga. Las vidas medias pueden ser del orden de 10^{-13} segundos.

Se esperaba que los estados encantados más ligeros fueran seis mesones, tres de ellos formados por combinación de un quark *c* con un antiquark $\bar{c}$, $\bar{d}$ o $\bar{s}$, y los otros tres por combinación de un antiquark $\bar{c}$ con un quark *u*, *d* o *s*. Los mesones que contienen un quark encantado se han llamado respectivamente D^+, D^0 y F^+; las antipartículas correspondientes son las D^-, $\bar{D}^0$ y F^-.

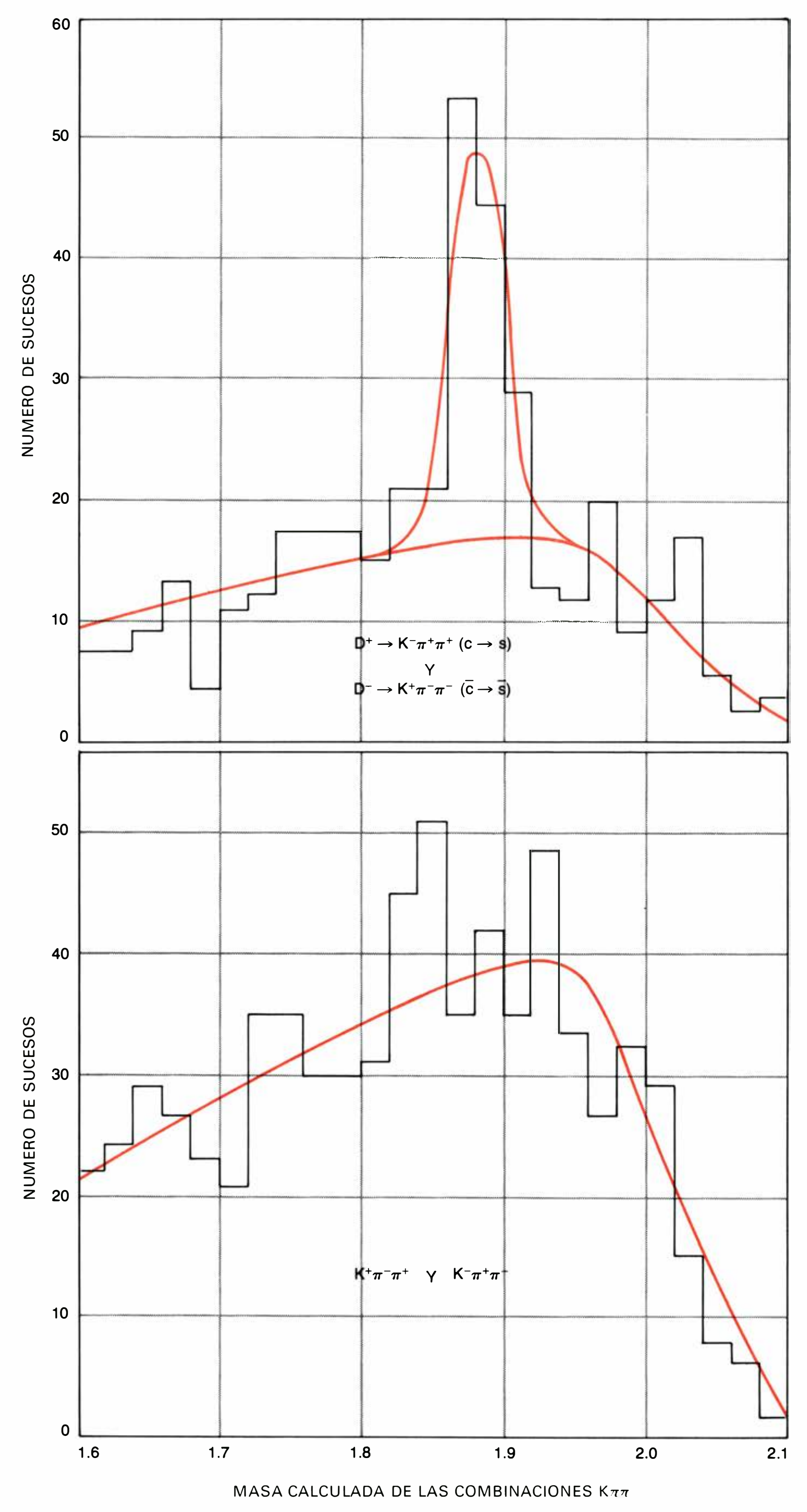
¿Cómo podría reconocerse a las partículas encantadas cuando se las encontrase? Aun cuando se descubrieran mesones de larga vida con las masas adecuadas, ¿cómo podría establecerse que su distintivo sería el encanto en vez de cualquier otra nueva propiedad de la materia? Las respuestas a esos interrogantes yacen en la especial relación que debe existir entre el quark encantado y el quark extraño para que pueda explicarse la supresión de las corrientes débiles neutras que cambian la extrañeza. En virtud de esa relación, las partículas encantadas que se desintegran débilmente deben incluir corrientemente entre sus productos de desintegración partículas extrañas. Los mesones *K*, las partículas extrañas de menor masa, hacen de signo distintivo de los sucesos que implican la desintegración de partículas encantadas.



LA PRIMERA PARTICULA ENCANTADA que se descubrió fue la D^0, compuesta de un quark *c* y un antiquark $\bar{u}$. Al contrario de la psi, la D^0 tiene encanto manifiesto: los números cuánticos del quark *c* no se cancelan por los de un antiquark $\bar{c}$ en el mismo hadrón. Por ser el mesón encantado más ligero, el D^0 puede desintegrarse sólo mediante la interacción débil, que puede transformar el quark encantado en un quark de otro tipo. A causa de la relación especial que existe entre extrañeza y encanto, el quark encantado se convierte corrientemente en un quark "extraño"; por consiguiente, se esperan hadrones extraños (como los mesones *K*) entre los productos de la desintegración. La D^0 se descubrió como un pico en la distribución de energía de mesones *K* y piones que tienen carga eléctrica total cero ($K^+\pi^-$ o $K^-\pi^+$). Tales combinaciones pueden producirse por muchos procesos que no se hallan relacionados, pero su número debe decrecer lentamente cuanto aumenta la masa del par. La presencia de un pico a una energía dada indica que una partícula dotada con aquella masa se encuentra efectivamente, desintegrándose en la combinación $K\pi$.

Estas propiedades de la desintegración de las partículas encantadas ponen de manifiesto un método experimental que permite hallar el encanto. Este método consiste en analizar las colisiones electrón-positrón por encima de la masa de la psi prima, buscando sucesos de muchas partículas (que corrientemente contienen hadrones) en los que haya partículas de larga vida, que incluyan mesones *K* entre sus productos de desintegración.

El proceso habitual seguido en tales investigaciones consiste en suponer un modo de desintegración determinado y seleccionar luego, de un gran conjunto de sucesos, todos aquellos que pueden ser ejemplos del modo supuesto. Por ejemplo, una hipótesis razonable es que algunas partículas encantadas puedan desintegrarse algunas veces en un mesón *K* y dos piones, y así se seleccionarían todos los sucesos que incluyen estas tres partículas. A partir de las mediciones del momento de los productos de desintegración, en cada suceso, se puede calcular la masa de la partícula padre. Si las partículas detectadas son realmente los productos de un determinado modo de desintegración, entonces la mayoría de las masas calculadas se agruparán en torno a cierto valor. Por otro lado, si las partículas seleccionadas están presentes en el suceso por pura casualidad, las masas calculadas tendrán una distribución amplia y sin estructura. Este método, muy corriente en la física



EL MESON D^+ constituye la prueba más fuerte de que las nuevas partículas observadas tienen la propiedad del encanto. El quark encantado en la D^+ debe desintegrarse en un quark extraño s, que es un constituyente del mesón K^-. La antipartícula de la D^+, la D^-, contiene un antiquark encantado, que debe originar un antiquark extraño, $\bar{s}$, el cual se encuentra en el mesón K^+. Las combinaciones de mesones K y piones coherentes con este modo de desintegración *(arriba)* muestran un pico a una energía cercana a 1,87 GeV. Es importante que las combinaciones que procediesen de la transición prohibida del quark encantado al antiquark extraño no muestren aumento a esta energía *(abajo)*.

de alta energía, se denomina "búsqueda de picos" ("bump hunting").

Nuestros primeros intentos de buscar picos a energías de aniquilación de unos 4 GeV no tuvieron éxito. La ausencia de una señal apreciable resultaba desalentadora, pero no era fatal para la hipótesis del encanto. Los mesones encantados podían desintegrarse de múltiples y distintos modos, de suerte que ninguno de éstos esté representado por un gran número de sucesos.

Entre tanto, empezaron a aparecer pruebas circunstanciales de la producción de partículas encantadas en experimentos que usaban técnicas muy diferentes. En tales experimentos, los sucesos se iniciaban por colisiones de neutrinos de alta energía sobre materia nuclear; puesto que las interacciones de los neutrinos se deben necesariamente a la fuerza débil, las partículas encantadas podían crearse individualmente en vez de crearse a pares. En Brookhaven y en la Organisation Européenne pour la Recherche Nucléaire (CERN) unas cuantas fotografías de cámara de burbujas mostraron sucesos que podían interpretarse como desintegraciones de partículas encantadas. En el Fermi National Accelerator Laboratory (Fermilab), un sistema de contadores de partículas registró numerosos sucesos candidatos, distinguibles por la presencia de dos leptones en el estado final. Los pares de leptones en el estado final constituyeron una prueba importante, aunque no del todo concluyente, de la desintegración de una partícula encantada de una masa próxima a 2 GeV. Con la excepción de una fotografía de cámara de burbujas, obtenida en Brookhaven, ningún suceso de éstos proporcionaba suficiente información para determinar con exactitud la masa de la partícula.

En el SPEAR, continuamos acumulando datos mientras estudiábamos las anchas estructuras de R a energías próximas a 4 GeV. En la primavera de 1976, con técnicas mejoradas para identificar partículas extrañas y con un conjunto de sucesos bastante más nutrido, obtuvimos, finalmente, un pico. Gerson Goldhaber y François Pierre, del Lawrence Berkeley Laboratory, encontraron picos pequeños, aunque significativos, a una energía de 1,863 GeV en dos modos de desintegración que producían mesones K y piones.

Una clase de sucesos tenía en el estado final un solo mesón K y un solo pión con cargas opuestas, es decir, $K^+\pi^-$ o $K^-\pi^+$. La otra clase de sucesos tenía un mesón K acompañado de tres piones con

cargas $K^+\pi^-\pi^+\pi^-$ o $K^-\pi^+\pi^+\pi^-$. Todas estas combinaciones de partículas tienen carga eléctrica total nula, y todas parecen provenir de la desintegración del mismo objeto. El candidato más plausible para este objeto es el mesón eléctricamente neutro D^o, formado por un quark c y un antiquark $\bar{u}$.

Desde entonces se han descubierto otros picos. El más importante de ellos proviene de la desintegración del compañero cargado del D^0, el mesón D^+. La prueba que da cuenta de este estado consiste en un aumento, a la energía de 1,868 GeV, en el número de sucesos que forman la combinación de partículas $K^-\pi^+\pi^+$. Significativamente, el modo de desintegración $K^+\pi^-\pi^+$, que tiene la misma carga total +1, no presenta el pico correspondiente a la misma energía. Esta sutil distinción entre estados de carga tiene una explicación simple y elegante en la teoría del encanto, y constituye la prueba más clara de la existencia del encanto. El mesón D^+ consta de un quark c ligado a un antiquark $\bar{d}$. La relación entre el encanto y la extrañeza impone que un quark c pueda desintegrarse mediante la interacción débil en un quark extraño s, pero no pueda convertirse en un antiquark $\bar{s}$. El mesón K^- contiene un quark s, y puede producirse, por tanto, en la desintegración del D^+; el mesón K^+, por otro lado, contiene un antiquark $\bar{s}$, y como exige la teoría no se observa en las desintegraciones del D^+. Ninguno de los modelos que pretenden explicar las nuevas partículas ofrece predicciones tan definidas y ninguno es tan coherente con los datos experimentales.

En las aniquilaciones electrón-positrón las partículas encantadas se crean invariablemente a pares, pero los picos por los que se las identifica representan los productos de desintegración de sólo un miembro del par. No obstante, se puede deducir alguna de las propiedades del elemento no identificado. Debido a que la partícula detectada está retrocediendo respecto de la que no ha sido detectada, ambos estados deben tener el mismo momento. A partir del momento y la energía total conocidos del sistema de partículas (que es igual a la energía de aniquilación del electrón y del positrón) puede determinarse la masa en reposo de la partícula que ha escapado.

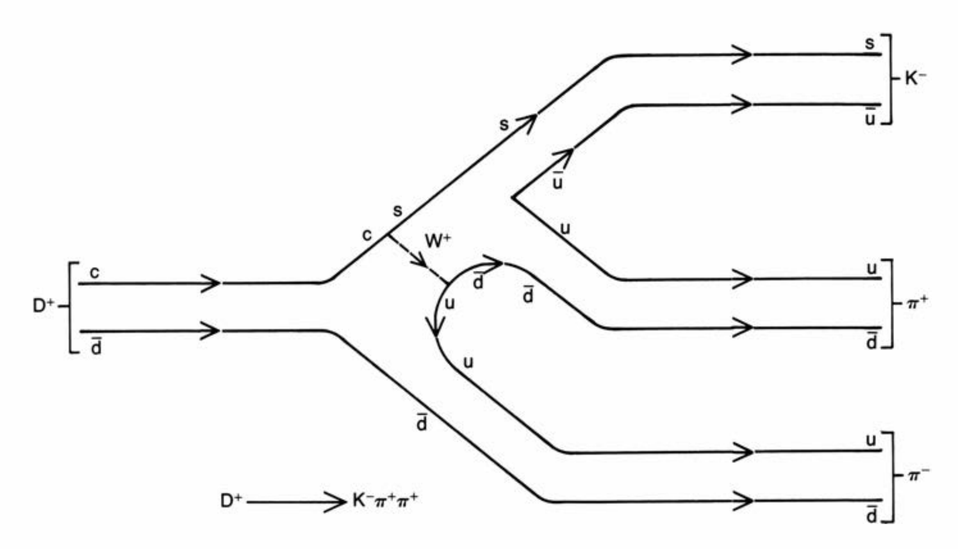
Cuando se analizaron así los sucesos en que se implicaban los mesones D^0 y D^+, encontramos que, en algunos casos, los objetos que retrocedían eran simples antipartículas $\bar{D}^o$ y D^-, con masas idénticas a las de las partículas detectadas, unos 1,87 GeV. Un pico más prominente en el espectro de masas de la partícula que retrocedía se observó cerca de 2,01 GeV. El pico se interpreta como indicación de estados excitados de los mesones, llamados D^{*0} y D^{*+}. Si esta interpretación es correcta, los estados de menor energía (D^0 y D^+) están formados por quarks con sus espines orientados antiparalelamente, de manera que el mesón tiene momento angular cero, mientras que en los estados excitados los espines de los quarks paralelos dan a los mesones una unidad de momento angular. De los espines dependen varias correlaciones angulares entre las partículas y sus productos de desintegración, de forma que puede comprobarse esta hipótesis. Las mediciones preliminares de esas correlaciones concuerdan con el modelo.

Una vez establecida la masa de la psi, pudieron producirse a voluntad grandes cantidades de estas partículas. Pero los éxitos experimentales más recientes no se han logrado tan fácilmente. Los estados de la materia con encanto manifiesto se han mostrado muy reacios a la observación. Después de tres años de búsqueda intensiva sólo se han registrado unos pocos centenares de sucesos en los que se han logrado identificar todos los productos de desintegración de una partícula encantada. De ellos, apenas un puñado corresponden a desintegraciones de mesones encantados. Debe haber también una extensa familia de bariones encantados, formados por tres quarks. En el Fermilab se han registrado unos cuantos sucesos que presentan visos de bariones encantados, pero está todavía por determinarse la mayoría de las propiedades de estas partículas.

La partícula encantada más elusiva es otro mesón, el F^+. Consta de un quark c y de un antiquark $\bar{s}$; por tanto, es a la vez encantado y extraño. Se están realizando varias búsquedas del F^+. Aunque no se dispone de pruebas concluyentes de su existencia, nadie duda seriamente de su futuro descubrimiento.

Aunque no se haya llegado a recabar todas las pruebas que confirmen la existencia del encanto, no puede dudarse del veredicto. El quark encantado empezó como un principio de simetría abstracto que relacionaba los quarks y los leptones, un principio que casi podía reducirse a un deseo de orden. En un principio, se propuso la hipótesis sin el soporte de ninguna observación experimental; hoy día muchos experimentos llevan a una conclusión que es cualquier cosa menos indiscutible. Un espectro entero de partículas testifica la existencia del quark encantado: la psi, los otros estados del charmonium y los mesones encantados D^0 y D^+. La prueba de más peso es también la más sutil: la relación especial entre extrañeza y encanto que se manifiesta en la desintegración de la D^+ en los tres mesones $K^-\pi^+\pi^+$ pero no en los $K^+\pi^-\pi^+$. Si esto no es encanto, se trata de una excelente copia.

Quizás el resultado más significativo del descubrimiento del encanto sea una mayor comprensión de la materia ordinaria. Se ha dicho a menudo que la física de altas energías necesita un equivalente al átomo de hidrógeno, que es el sistema comparativamente simple de

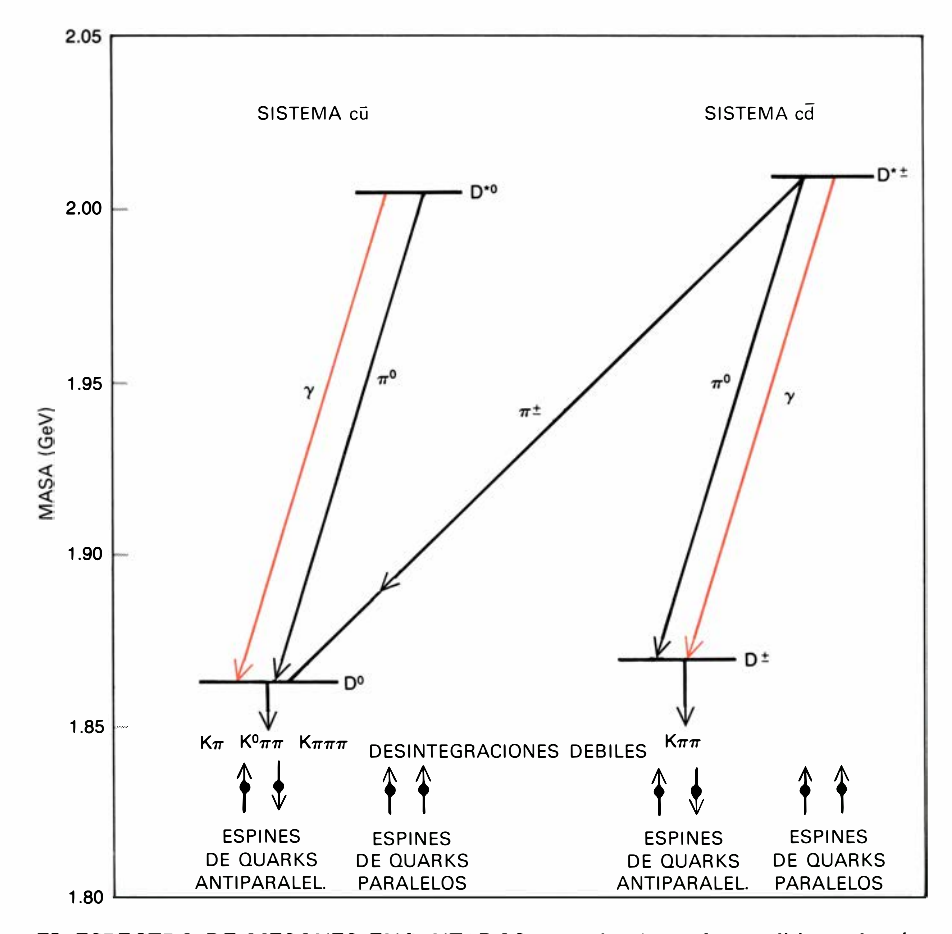


LA DESINTEGRACION DEL MESON D^+ tiene lugar a través de una interacción débil entre quarks. Mientras que la fuerza fuerte sólo puede crear o aniquilar pares de quarks y antiquarks de la misma clase, la fuerza débil puede cambiar una clase de quark en otra. En la desintegración de la D^+ la fuerza débil se transmite por una partícula llamada W^+. El quark c emite un W^+ y se transforma en un quark s; el W^+ entonces se descompone en un quark u y un antiquark $\bar{d}$. El análisis de la interacción muestra que no se conserva el encanto ni la extrañeza, aunque sí la carga eléctrica.

partículas ligadas que sirvió de constante punto de referencia para el desarrollo de la mecánica cuántica hace 50 años. En cierto sentido, el modelo de quarks trata todos los hadrones como sistemas simples parecidos a los átomos, pero esta estructura nunca se vio muy claramente en los experimentos hasta el descubrimiento de la partícula psi. Ahora las agudas transiciones entre los estados del charmonium, con emisión de fotones, proporcionan una analogía exacta a la espectroscopía atómica, pero a energías un millón de veces superiores. El descubrimiento del encanto marca el inicio de la química de los quarks.

La existencia de los quarks encantados puede haberse probado, pero sería totalmente erróneo dar por concluida la cuestión. Un importante problema es el valor de la razón *R*. El modelo de los quarks más plausible predice un valor de 3,3, pero la razón observada se halla cercana a 5,5. Una vez más, *R* parece ser demasiado grande. Además, hay un número creciente de sucesos anómalos, observados tanto en SPEAR como en DORIS, que no se explican fácilmente por las versiones actuales del modelo de los quarks, aun con inclusión del encanto. Estos sucesos se caracterizan por la presencia de dos partículas cargadas, a menudo leptones, y partículas neutras no detectadas, que pueden ser neutrinos.

Actualmente, la hipótesis que se muestra más capacitada para dar cuenta de los sucesos anómalos y de casi todo el exceso de *R* es la que propugna la aparición de un quinto miembro de la familia de leptones. Tendría una masa de unos 2 GeV (o 4000 veces la masa del electrón) y podría estar acompañado por aun otro nuevo leptón, un neutrino. Desde luego, si hay leptones de más, la atractiva simetría entre leptones y quarks quedaría destruida. La solución obvia sería postular más quarks, y ya se ha informado de pruebas preliminares de que hay partículas dotadas de una quinta clase de quark, en interacciones de neutrinos de alta energía en Fermilab. Se necesitarán más experimentos para confirmar tal descubrimiento, pero el futuro de la física de partículas elementales parece presentarse al menos tan brillante como lo ha sido en el pasado inmediato.



EL ESPECTRO DE MESONES ENCANTADOS que se ha observado en colisiones electrón-positrón comprende cuatro estados y sus antipartículas. Los estados de menor masa tienen los espines de los quarks antiparalelos y espín total cero; los estados excitados D^* constan de los mismos quarks, pero tienen los espines de los quarks paralelos, dando a los mesones un espín total 1. El mesón encantado que falta, que tendría también estados excitados, es el F^+. Este mesón está formado de un quark c y un antiquark $\bar{s}$; se trata, pues, de una partícula con encanto y extrañeza. Hay indicios de que se haya visto el F^+ en un experimento europeo, pero la prueba no es concluyente.

Nota complementaria

Con posterioridad a la redacción de la versión americana de este artículo, se han realizado algunos descubrimientos importantes que vienen a confirmar los puntos de vista expuestos.

El primero de ellos ha sido la confirmación de la existencia del mesón F. Un grupo experimental que trabaja en el DORIS ha encontrado la partícula F (de espín cero) y la F^* (la excitación de espín 1) con unas masas de 2,03 y 2,14 GeV, respectivamente. Según el modelo de quarks, el contenido en quarks del mesón F^+ es $c\bar{s}$ y, dada la especial facilidad del quark c para convertirse, mediante la interacción débil, en el quark s, es de esperar que el mesón F se desintegre en sistemas que tengan una importante contribución de la combinación $s\bar{s}$; un tal sistema lo constituye el mesón η, por lo que es de esperar que uno de los modos preferidos de desintegración del F^+ sea dando $\eta\pi^+$. Si además, como sucede con el mesón D, existe un estado excitado F^*, que se produciría a la vez que el F, de masa próxima a la del F y que se desintegraría en éste emitiendo un fotón, lo que se debe buscar son sucesos en que aparezca un mesón η acompañado de un fotón de baja energía. Es en esta búsqueda donde han aparecido los mesones F con las masas antes citadas.

Este descubrimiento, junto con algunos nuevos estados del charmonium ya previstos y que se han encontrado, y con el hallazgo de algún nuevo barión encantado, tiende a completar y dar consistencia al modelo de los quarks.

Otro hecho importante ha sido el descubrimiento de, al menos, un leptón pesado. La existencia de leptones pesados era esperada desde hace unos pocos años a partir de argumentos puramente teóricos. A resultas de las aniquilaciones electrón-positrón se han encontrado sucesos que podrían interpretarse como la transformación del par electrón-positrón en un fotón, que, a su vez, crea el leptón pesado, llamado tau (τ) y su antipartícula. La masa del nuevo leptón, que se añade a la familia del electrón y el muón, sería de 1,9 GeV; probablemente, tendría asociado un neutrino de masa cero. La situación experimental no es todo lo clara que se desearía en cuanto a la determinación de las características del tau; sí es claro, sin embargo, que en todos los experimentos en que debería verse una señal del tau, dicha señal se manifiesta.

Lo que nos importa aquí de la existencia de nuevos leptones es que tal existencia hace muy atractiva la presencia de nuevas clases de quarks; en este contexto debe enmarcarse el descubrimiento de una nueva partícula de masa casi 20.000 veces la masa del electrón. A finales de junio, el grupo experimental formado por físicos de la Universidad de Columbia, Fermilab y Stony Brook, bajo la dirección de Leon Lederman, anunció el descubrimiento de esta nueva partícula a la que llamaron upsilon (υ).

La evidencia de la upsilon apareció estudiando colisiones de protones producidos por el acelerador de Fermilab sobre materia nuclear. Se buscaban sucesos en que se produjesen pares leptón-antileptón, concretamente muón negativo-muón positivo; la medición de la energía del par nos dará, si se presenta una acumulación de sucesos a ciertas energías, la masa de la partícula madre que ha producido el par. Si esta partícula no existe, es decir, si la producción del par ha sido un suceso aleatorio, no habrá acumulación de sucesos a ninguna energía. En el experimento se encontraron los picos correspondientes a los estados del charmonium, pero también se halló una clara acumulación de sucesos a energías próximas a 9,5 GeV, lo que probaba la existencia de la upsilon. De hecho, los últimos resultados indican que no se trata de un solo pico, sino que hay, al menos, dos a los que ya se conoce como upsilon, de masa 9,4 GeV y upsilon prima, de masa 10 GeV.

Todavía es pronto para decidir en qué consiste esta nueva partícula, aunque se tiende a establecer una analogía con lo ocurrido con el charmonium. El causante sería ahora una nueva clase de quarks llamada "belleza" ("beauty") o "verdad" ("truth"). De esta manera, las partículas upsilon serían estados ligados de los nuevos quark-antiquark. Sin embargo, las condiciones experimentales en esta nueva zona de energías son peores y los anillos de colisión de electrones actuales no permiten investigar en esta zona; habrá que esperar, pues, para obtener unos buenos datos experimentales, a la puesta en marcha de los anillos PETRA en DESY y PEP en SLAC, que se encuentran en avanzada fase de realización.

(Ramón Pascual, catedrático de Física Teórica, de la Universidad Autónoma de Barcelona.)

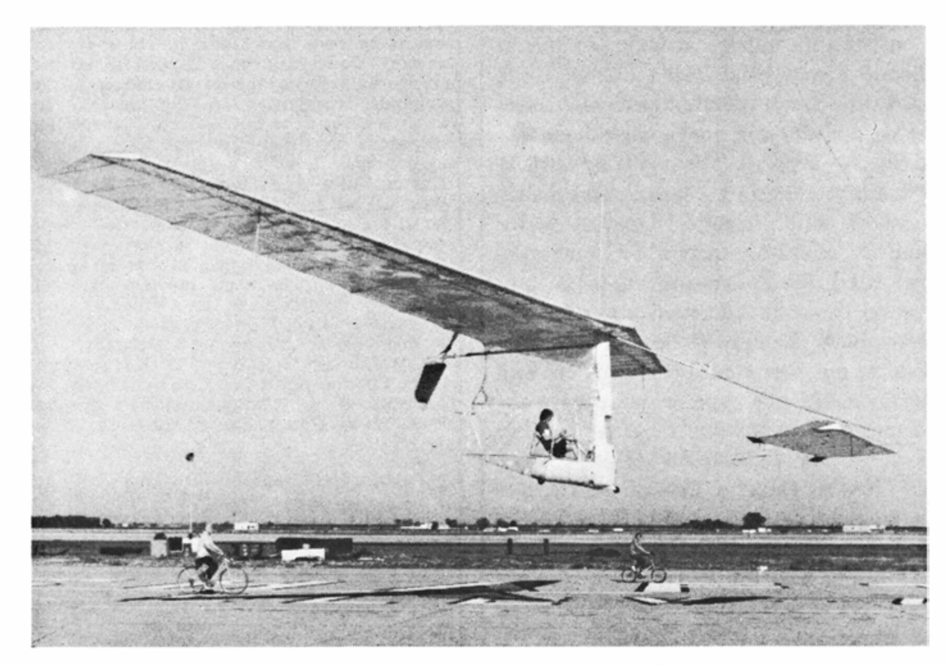
Ciencia y sociedad

Alas de telaraña

Que un hombre vuele como un pájaro, propulsado por su propia potencia, es el viejo sueño de la aeronáutica, y se remonta por lo menos al mito griego de Dédalo, el inventor, quien escapó del Laberinto en unas alas cubiertas de plumas sujetas con cera. La era moderna del vuelo a propulsión humana comenzó en 1935, cuando dos ingenieros ingleses crearon una "bicicleta volante" impulsada por hélice que se desplazaba breve trecho por el aire. En los años cincuenta, Henry Kremer, industrial inglés, ofreció, bajo los auspicios de la Real Sociedad Aeronáutica, un substancioso premio (que vale ahora unos 86.000 dólares) para el primer vuelo a propulsión humana que cumpliera ciertas difíciles condiciones: el vehículo debía despegar sin ayuda desde terreno llano, con un viento de 16 kilómetros por hora o de menor velocidad, volar haciendo un ocho en torno a dos torres distantes 800 metros entre sí y pasar sobre una valla de tres metros de altura a la partida y a la llegada. Se prohibía el uso de potencia acumulada o de gases más ligeros que el aire, como también toda detención o ayuda a lo largo del recorrido.

A lo largo de los años, el desafío planteado por Kremer dio origen a una serie de máquinas voladoras delicadas y de anchas alas. Varias pudieron efectuar prolongados vuelos rectos, pero ninguna pudo navegar siguiendo el circuito; los giros resultaban ser demasiado difíciles, tanto en lo concerniente a la maniobrabilidad como en cuanto a la potencia. Ahora, un nuevo tipo de vehículo aéreo creado por Paul B. MacCready, presidente de AeroVironment, Inc. de Pasadena, California, ha salido finalmente victorioso donde otros fracasaron.

El ingenio de MacCready, el *Gossamer Condor* (cóndor telaraña), lleva un nombre apropiado: mide 9,144 metros de longitud, 2,438 metros de altura, 29,26 metros de envergadura y su área total de sustentación es de 77,57 metros cuadrados; sin embargo, pesa sólo 31,75 kilogramos. El desproporcionado tamaño del ala es necesario para lograr la máxima sustentación a poca velocidad. El vehículo se propulsa pedaleando una cadena de bicicleta conectada a una hélice situada tras la cabina, que cuelga del centro del ala. Un plano fijo horizontal montado en un fino larguero que se prolonga hasta la parte frontal del ala, y provisto de dos pequeños alerones, ayuda a controlar el cabeceo y la guiñada y evita la necesidad de una superficie de control vertical. El vehículo se maniobra para girar torciendo directamente el ala con ayuda de una palanca, que puede fijarse en un radio de giro predeterminado. Una vez fija la torsión del ala, sigue siendo posible manipular el plano estabilizador para mantener el control durante el giro.



Gossamer Condor *planeando, en el primer vuelo a propulsión humana*

La principal diferencia entre el *Gossamer Condor* y los anteriores artefactos voladores a propulsión humana consiste en que fue inspirado por la tecnología del planeador colgante. Estos livianos planeadores dieron a MacCready la idea de usar un esqueleto de tubos de aluminio atirantado con cuerda de piano. Escribe MacCready: "El hecho de que la estructura básica fuese fácil de construir, modificar y reparar hizo posible un programa integrado de desarrollo. Podíamos construir una nueva ala en dos semanas y un nuevo estabilizador en tres días". "El diseño del vehículo", dice, "es una mezcla de las tecnologías aeronáuticas más sencilla y más adelantada". Aunque la superficie aerodinámica y la hélice fueron proyectadas con ayuda de un ordenador por el vicepresidente de AeroVironment, Peter P. S. Lissaman, jamás se dibujaron planos del vehículo. El diseño evolucionó en cambio a lo largo de un año, durante el cual se construyeron y perfeccionaron 12 modelos sucesivos.

No se precisó ningún material hecho de encargo. El fuselaje se construyó de madera de balsa, cartón ondulado (que también forma el borde de ataque del ala), láminas de Styrofoam y una delgada cubierta de mylar. Obviamente, se intentó por todos los medios minimizar el peso. El proyecto exigió una inversión de 25.000 dólares y miles de horas-hombre no retribuidas. MacCready y sus colaboradores construyeron primero modelos en pequeña escala para verificar las técnicas de arriostramiento y de construcción. Pasaron luego a una versión de 26,82 metros de envergadura pero sin hélice. Finalmente, construyeron una versión a tamaño natural con una hélice y un asiento de bicicleta. Para pilotar el aparato, MacCready contrató a Bryan Allen, ciclista de competiciones, de 24 años de edad, que es también un experto piloto de planeador colgante.

El artefacto fue sometido a prueba en el desierto de Mojave. En esa ocasión se descubrió que los extremos de la alas eran demasiado anchos para permitir una maniobrabilidad razonable a la baja velocidad del aparato (de unos 11 a 18 kilómetros por hora). Cuando se hicieron más angostos los extremos de las alas, se hizo necesario disminuir aún más la resistencia aerodinámica encerrando al piloto en una cabina fuselada de mylar. Con el piloto en su sitio, el aparato terminado pesa 94 kilogramos y necesita alrededor de 0,35 caballos de vapor para volar en línea recta. Dice MacCready: "Es el avión absolutamente mínimo".

En febrero, MacCready y sus colaboradores trasladaron el avión al aeropuerto del Condado de Kern, en Shafter, California. Se eligió ese aeropuerto porque está relativamente abandonado, tiene un gran hangar y los vientos que allí soplan son más leves que en la mayoría de los aeródromos de California meridional. El 13 de agosto, después de algunas modificaciones finales del fuselaje, MacCready decidió intentar ganar el premio Kremer. Su confianza se vio reforzada por una "confluencia de factores benéficos": Allen estaba entrenado para dar lo mejor de sí, las condiciones de viento eran satisfactorias y Bob Richardson, del Departamento de Aeropuertos del Condado de Kern, que había sido aprobado por la Real Sociedad Aeronáutica para verificar un vuelo satisfactorio, estaba presente.

Conforme Allen pedaleaba vigorosamente, el *Gossamer Condor* aceleró por la pista a lo largo de 9 metros y despegó lentamente. Moviéndose a unos 18 kilómetros por hora, siguió el recorrido especificado en forma de ocho en torno a dos hitos distantes 800 metros entre sí, aunque a causa de la gran envergadura y escasa velocidad del aparato los giros hubieron de ser amplios y planos. Al terminar con éxito el vuelo, el aparato había recorrido, entre despegue y aterrizaje, dos kilómetros en siete minutos y 22,5 segundos, y había ganado para MacCready y sus colaboradores el premio Kremer.

Ahora que ha pasado la competición, MacCready quiere descansar de una actividad que le ha ocupado casi todas las horas de su tiempo libre durante el pasado año. El *Gossamer Condor,* dice, era "un vehículo burdo con sólo una finalidad: ganar el premio Kremer. Fue una solución intermedia entre un diseño sofisticado y una cantidad finita de tiempo, dinero y esfuerzo". Pero él cree que su diseño básico podría afinarse para disminuir el peso de 31,75 a 25 kilógramos y hacer más suave la superficie alar. El aparato sólo necesitaría entonces 0,29 caballos de vapor para mantenerse en el aire, permitiendo que un atleta de la bicicleta se mantenga volando en él varias horas. En cuanto al aparato victorioso, es candidato al retiro. MacCready espera que encuentre un lugar en la Smithsonian Institution.

Claudio Ptolomeo: impostor

A Claudio Ptolomeo, erudito griego que floreció en Alejandría en el siglo II, se le ha considerado desde hace largo tiempo como el mayor astrónomo de la antigüedad. A su tratado astronómico, *'E Mathematike Syntaxis,* escrito en 142 d.C., se le ha llamado la obra de astronomía matemática más importante hasta que Copérnico publicó su *De Revolutionibus Orbium Coelestium,* 1400 años más tarde. En la *Syntaxis,* Ptolomeo expuso un sistema matemático para predecir las posiciones futuras de los planetas, en el cual la tierra está en reposo y el sol, la luna, los planetas y las estrellas se mueven en su entorno en epiciclos circulares. La teoría planetaria ptolemaica influyó de tal manera en el desarrollo de la astronomía islámica que los astrónomos árabes dieron a la *Syntaxis* el nombre de Almagesto, es decir, "el más grande". La teoría ptolemaica moldeó asimismo el desarrollo de la cosmología europea hasta que se impuso la labor de Copérnico y de Kepler. Más recientemente, la *Syntaxis* ha sido importante tanto para los historiadores como para los astrónomos, porque Ptolomeo citó teorías y observaciones astronómicas que atribuyó a sus predecesores griegos y babilonios; en ciertos casos, su libro es la única fuente de información que sobrevive acerca de los inicios de la cultura griega y de la cronología babilónica. Ultimamente, según Robert R. Newton, de la Johns Hopkins University, en su libro *The Crime of Claudius Ptolomeus,* que publicará en breve la Johns Hopkins University Press, Ptolomeo se revela como el impostor de mayor éxito de toda la historia de la ciencia.

La prueba principal del fraude se halla en la propia *Syntaxis.* De acuerdo con análisis estadísticos de Newton, la exactitud de ciertas observaciones de los planetas y estrellas que Ptolomeo pretendía haber realizado es tan grande que en algunos casos las probabilidades son de una en mil millones respecto de que las haya hecho con los instrumentos que describe. En otros casos, las observaciones de equinoccios y solsticios que pretendía haber realizado Ptolomeo están desfasadas en más de un día, error tan irrazonablemente craso que es diez veces mayor que el error en que incurrieron sus predecesores en observaciones semejantes. En la versión del catálogo de estrellas de su predecesor Hiparco, "puesta al día" por Ptolomeo, las posiciones de unas 1000 estrellas que éste alegaba haber observado se obtuvieron desplazando cada posición en una magnitud igual. En muchos ejemplos, las conclusiones de Ptolomeo basadas en las observaciones que alega haber realizado son minuciosas hasta el último dígito, aun cuando sus cálculos errasen. Todas estas rarezas han inducido a Newton a concluir que Ptolomeo dedujo sus observaciones de sus tablas, en vez de obrar a la inversa. Escribe Newton: "Elaboró ciertas teorías astronómicas y descubrió que no eran compatibles con la observación. En vez de abandonar las teorías, fabricó deliberadamente observaciones a partir de las teorías, de modo que pudiesen alegar que las observaciones probaban la validez de sus teorías".

Newton ha descubierto asimismo que gran parte de las observaciones que Ptolomeo atribuyó a otros astrónomos no era documentación antigua que él hubiese rescatado sino datos que había falsificado. "A ojo de buen cubero", calcula Newton, "el percatarse del fraude de Ptolomeo destruye la mitad de lo que hemos estado aceptando como astronomía griega". Además, la falsificación de Ptolomeo quizá haya llegado hasta inventar la duración de los reinos de los reyes de Babilonia. Como gran parte de la reconstrucción moderna de la cronología babilónica se ha basado en una lista de reyes que Ptolomeo usó para fijar las fechas de presuntas observaciones babilónicas, según Newton "toda la cronología pertinente debe revisarse ahora y debe eliminarse toda dependencia respecto de la lista de reyes de Ptolomeo".

Newton ha examinado la posibilidad de que un ayudante de Ptolomeo proporcionara a éste observaciones falsas por pereza, pero cree improbable el que ese ayudante hubiese falsificado también la historia. Newton menciona también como una vaga posibilidad el que la *Syntaxis* no hubiese sido escrita por Ptolomeo. Newton no ha encontrado aún prueba ninguna sobre la motivación del fraude, aunque sugiere que la más probable sea simplemente que Ptolomeo quería ser conocido como un gran astrónomo. "Quizá haya visto, al iniciar su carrera, que no reunía las condiciones, de modo que recurrió a la única otra manera de satisfacer su ambición, que consistió en reemplazar la capacidad con el fraude".

Para Newton, el problema más difícil que presenta la *Syntaxis* es: ¿Por qué ha sido aceptada como una gran obra durante 18 siglos? Durante ese tiempo, algunos astrónomos descubrieron ciertas contradicciones en ella, pero al parecer el único erudito que puso seriamente en tela de juicio el valor de la obra fue J. B. J. Delambre, astrónomo francés de principios del siglo XIX.

"Está claro", dice Newton al concluir su libro, "que ninguna afirmación hecha por Ptolomeo puede aceptarse a menos que esté confirmada por escritores totalmente independientes de él en el asunto de que se trate. Todas las investigaciones de historia o de astronomía que se hayan fundado en la *Syntaxis* deberán rehacerse en adelante."

Radar de imágenes laterales aerotransportado

Mediante una antena de radar montada en el fuselaje de un avión y enfocada hacia uno de sus lados se pueden obtener imágenes detalladas de la superficie terrestre, con independencia de las condiciones meteorológicas y la hora del día o la noche

Homer Jensen, L. C. Graham, Leonard J. Porcello y Emmett N. Leith

Las primeras imágenes aéreas de la superficie terrestre, obtenidas con fines topográficos o militares, consistieron en fotografías ordinarias tomadas desde globos a mediados del siglo XIX. El progresivo desarrollo de las cámaras, las películas y los aviones hizo posible que el campo de la fotografía aérea fuera cada vez más amplio. Hacia 1920, las fotografías verticales tomadas desde los aviones se convirtieron en la base de la fotogrametría aérea. Hoy en día, casi todos los mapas geológicos y topográficos están basados en imágenes de la superficie terrestre tomadas desde aviones o satélites artificiales.

En los años cincuenta se inventaron unos dispositivos de captación de imágenes cuya sensibilidad se extendía más allá de la región visible del espectro, particularmente a la región del infrarrojo. Mediante la utilización de sensores infrarrojos, la región del espectro electromagnético en la que puede observarse la superficie terrestre se extiende desde los 0,4 micrómetros (correspondiente a las longitudes de onda más cortas del espectro visible) hasta los 10 micrómetros (longitudes de onda del infrarrojo térmico). El funcionamiento de todos estos dispositivos se basa en la detección de energía. Esta energía es la luz solar reflejada por la superficie de la tierra y las estructuras levantadas por el hombre o la radiación que emiten estos cuerpos en función de su temperatura y emisividad naturales.

Sin embargo, para longitudes de onda visibles e infrarrojas, la atmósfera absorbe una cantidad significativa de radiación, incluso en el caso de que el tiempo esté despejado. Cuando el tiempo está nublado o lluvioso, el funcionamiento de los sensores resulta muy seriamente afectado. A pesar de que los observadores aéreos tienen a su disposición un intervalo espectral relativamente amplio, durante mucho tiempo no han podido satisfacer su deseo de contar con una especie de cámara mágica que pudiera operar desde el aire para obtener fotografías de la superficie terrestre, independientemente de la iluminación y las condiciones meteorológicas.

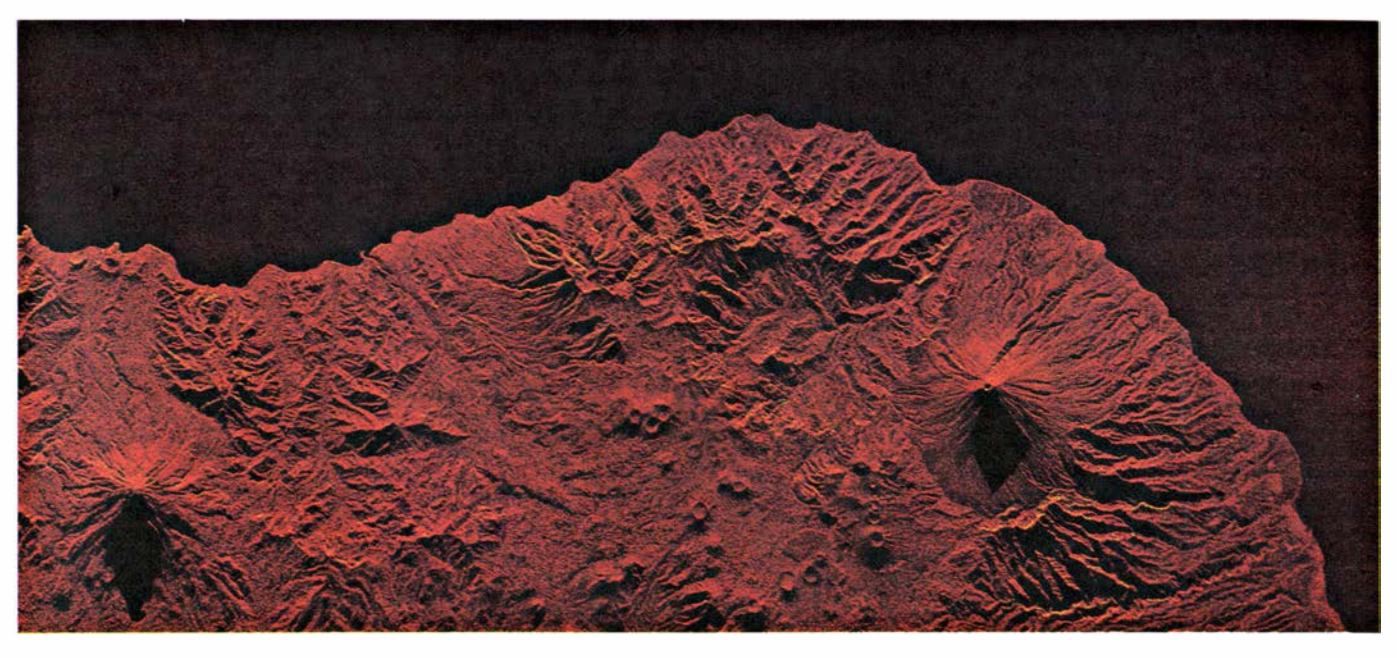
Los requisitos de un sistema de producción de imágenes para todo tiempo, sea de día o de noche, son, en principio, muy simples. Al objeto de que su funcionamiento sea independiente de la luz solar, el sistema necesita su propia fuente de iluminación. Con el fin de que pueda atravesar las nubes, el sistema debe utilizar un tipo de radiación que no sea atenuada o dispersada por el vapor de agua. Existe un dispositivo de producción de imágenes que cumple estos dos requisitos: un sistema de radar que trabaje en la región de las microondas, es decir, con longitudes de onda comprendidas entre uno y treinta centímetros.

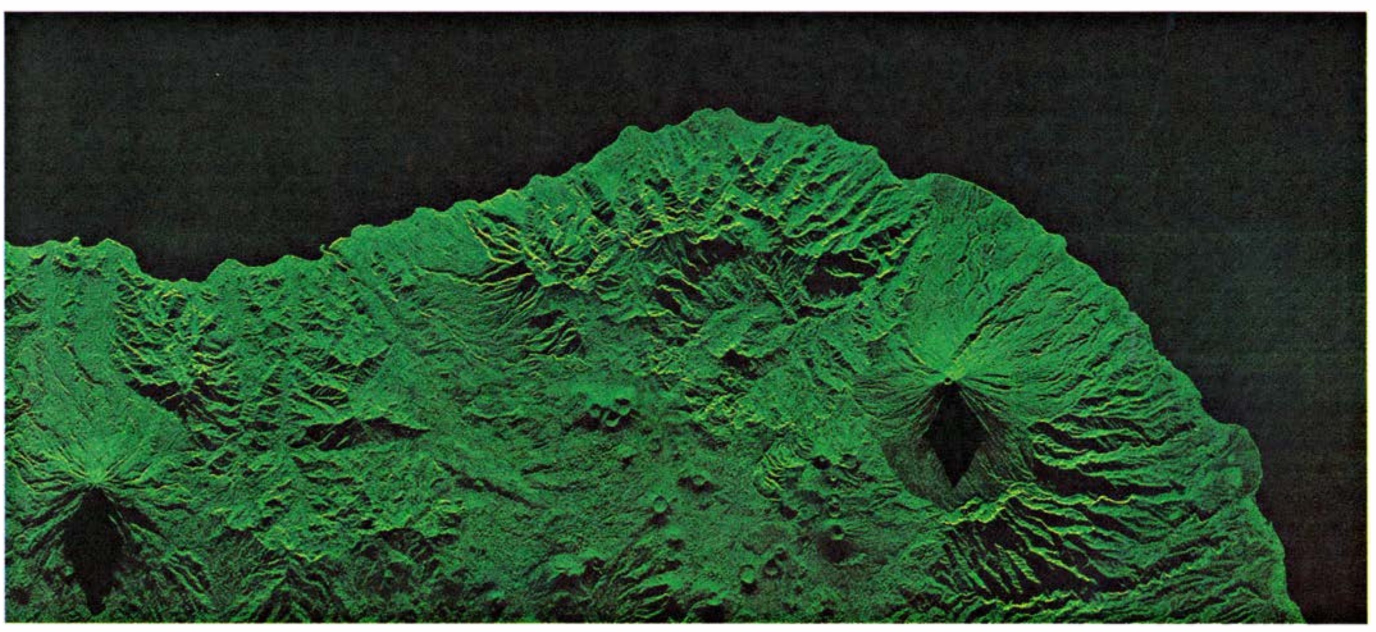
Los sistemas de radar proporcionan su propia iluminación y las microondas de ciertas frecuencias resultan poco afectadas por la atmósfera y las condiciones climatológicas. Sistemas de radar que trabajan con estas longitudes de onda son los que vigilan el clima desde el suelo y desde satélites, contribuyen a controlar el tráfico aéreo y han medido la distancia que nos separa de la luna y de los planetas. Aunque las longitudes de onda de las microondas son 100.000 veces mayores que las de la luz visible, todavía resultan lo suficientemente pequeñas como para permitir una resolución adecuada para la mayoría de los fines geológicos y geográficos. De hecho, las imágenes de la superficie terrestre que proporciona un sistema de radar que trabaje con las citadas longitudes de onda y que se transporte a bordo de un avión, se caracterizan por su extraordinario detalle y su espectacular sensación de relieve. En el presente artículo, nos limitaremos a estudiar los sistemas de radar aerotransportados que ofrecen imágenes de los terrenos cercanos a través de observaciones dirigidas hacia los lados del avión.

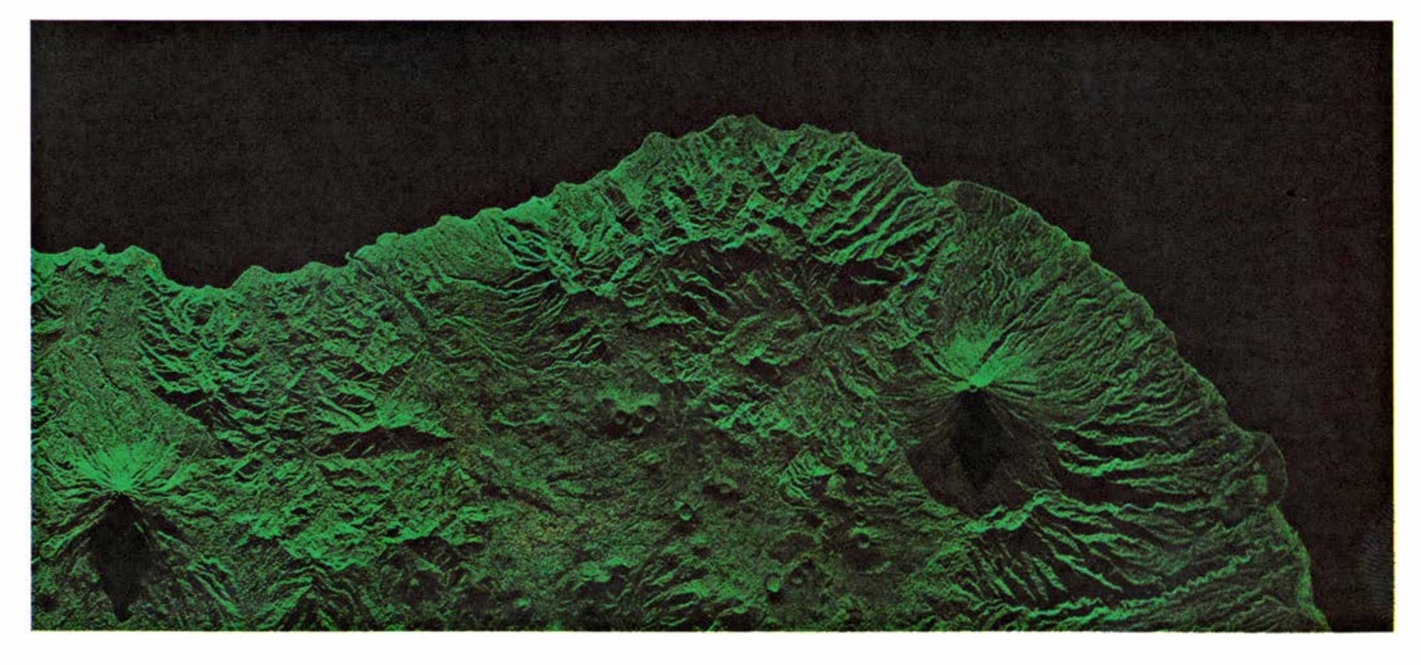
Resolución del sistema

Precisamente podemos empezar preguntándonos por qué la antena del radar está dirigida hacia uno de los lados del avión. Casi todo el mundo está familiarizado con los sistemas de radar de barrido circular, cuya imagen se obtiene mediante una antena giratoria. Las predicciones del tiempo que presentan algunas televisiones, por ejemplo, utilizan este sistema para poner de manifiesto la

ESTAS IMAGENES DE RADAR LATERAL corresponden a una isla de Indonesia y están caracterizadas por su gran sensación de relieve. De acuerdo con lo que indican las sombras, la ruta del avión que tomó estas imágenes era paralela a la dirección longitudinal de las mismas. En estas figuras se representan 40 kilómetros de costa, al sur de la isla de Flores. El pico de la izquierda es un volcán activo de unos 2000 metros de altura; el pico de la derecha es un volcán apagado de la misma altura. El ángulo rasante típico del radar lateral permite también poner de manifiesto, en el centro de las figuras, otros cráteres volcánicos más pequeños, así como los picos y los valles próximos al mar. Las tres fotografías se han obtenido a partir de una misma imagen de radar. Los colores de las fotografías son diferentes debido a que cada una de ellas se obtiene y se reproduce utilizando una longitud de onda diferente de luz monocromática. Sin embargo, en todas estas imágenes se pone de manifiesto más de un color, porque en ciertas zonas de la fotografía la luz monocromática era lo suficientemente intensa como para excitar una segunda capa de la emulsión en la película de color. Las imágenes en color pueden proporcionar más información que las de en blanco y negro.







intensidad de la precipitación en la zona cubierta por la correspondiente antena meteorológica. En un sistema de radar de barrido circular, el retículo del tubo de rayos catódicos gira sincrónicamente con la antena y la intensidad variable con que se refleja la energía viene dada por la variación del brillo del punto luminoso correspondiente. Como el tubo de rayos catódicos hace persistir la imagen de cada barrido, puede observarse una imagen continua.

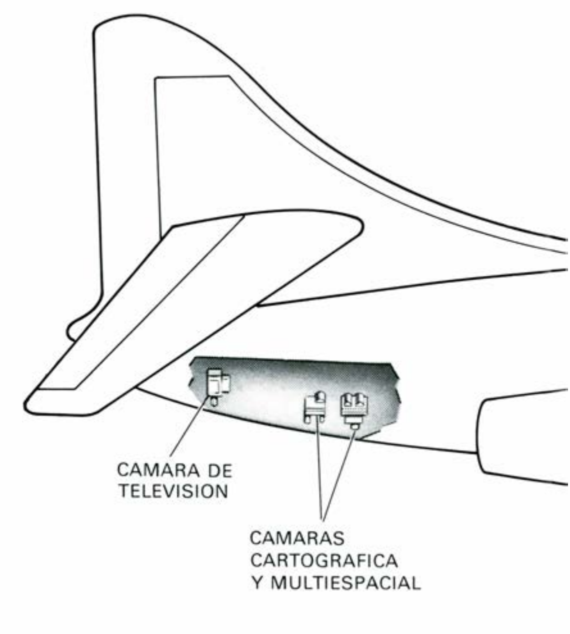
La experiencia ha demostrado que las imágenes que nos proporcionan los sistemas de radar de barrido circular transportados a bordo de aviones suelen caracterizarse por su pobre definición. Este hecho es una consecuencia de que la mayoría de las antenas de radar de barrido circular son bastante pequeñas; para obtener una buena resolución angular es necesario utilizar un sistema cuya apertura sea grande respecto a la longitud de onda de la radiación recibida. En otras palabras, la resolución de una lente o antena de gran apertura es mayor que la resolución de una lente o antena de poca apertura. Evidentemente, es difícil montar a bordo de un avión una antena giratoria de grandes dimensiones, pero en cambio resulta fácil instalar en su fuselaje una antena fija de cinco metros de longitud, por ejemplo, orientada hacia uno de los lados del aparato.

Las imágenes que proporciona un sistema de radar lateral aerotransportado, constituyen un registro gráfico de las propiedades de reflexión del terreno para longitudes de onda en la región de las microondas. Estas imágenes, conservadas sobre una película fotográfica, son muy similares a las de una fotografía ordinaria tomada desde el aire. Sin embargo, entre estos dos tipos de imágenes existen también ciertas diferencias fundamentales. Al contrario de lo que sucede en una fotografía aérea hecha a la luz del sol, la antena del radar es la fuente de iluminación del terreno y la imagen obtenida en este caso se basa en su propia energía reflejada. Puesto que las ondas de radar se mueven en línea recta, las regiones ocultas por colinas u otros objetos verticales no resultan "iluminadas" y, por lo tanto, ni reciben ni devuelven los impulsos de microondas. En la pantalla del radar, las sombras resultantes son totalmente oscuras. En cambio, las imágenes que se obtienen de fotografiar unos accidentes topográficos de este tipo, a la luz del amanecer o el atardecer, aparecen débilmente iluminadas por el efecto de la difusión óptica en la atmósfera. El carácter más o menos detallado de la imagen del radar viene determinado por la longitud de onda y la polarización de la señal incidente y por las propiedades eléctricas y geométricas de las superficies reflectoras existentes en el terreno observado. Vamos a considerar solamente la longitud de onda de la señal y las propiedades geométricas del terreno.

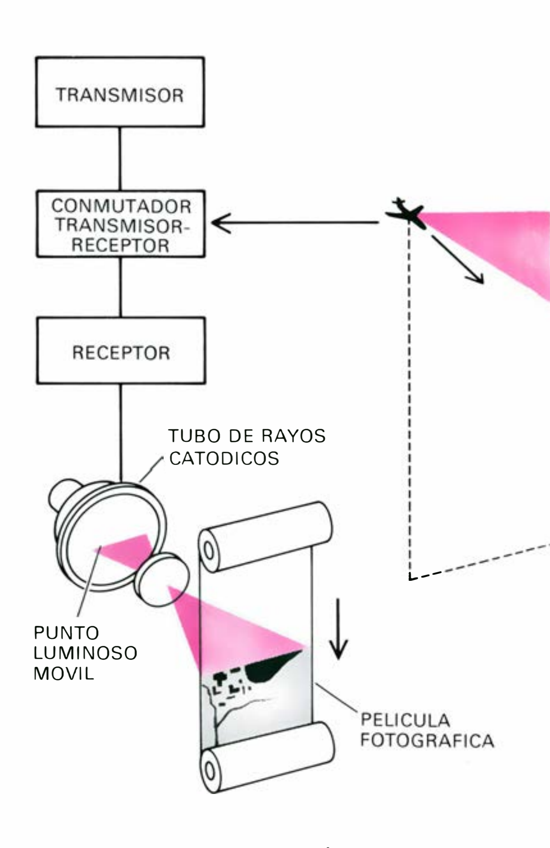
Geometría y reflexión

El hecho de que una superficie dada aparezca rugosa o lisa en la imagen del radar viene determinado por el valor de la longitud de la onda electromagnética en relación con el tamaño de los accidentes del terreno. Por ejemplo, una superficie que resulta rugosa a la longitud de onda visible de 0,5 micrómetros puede aparecer relativamente lisa (es decir, sin detalle de accidentes) a una longitud de onda de tres centímetros. Una superficie rugosa difunde la energía incidente en todas direcciones, devolviendo a la antena solamente una pequeña fracción de dicha energía. En cambio, una superficie lisa refleja la energía incidente en una dirección determinada, comportándose como un espejo. Si la superficie lisa es perpendicular al haz del radar incidente, la energía devuelta a la antena es intensa: en otro caso, el haz no vuelve a la antena. Algunas características del terreno, como por ejemplo, un campo de maíz, son reflectores difusos tanto para la región visible como para la región de las microondas del espectro. Otras características del terreno, verbigracia, una zona de estacionamiento de cemento, se muestran como reflectores difusos para la región visible y reflectores especulares para la región de las microondas. En conclusión, existen más reflectores especulares en relación con la captación de imágenes con microondas que los que se ponen de manifiesto en las fotografías realizadas con luz visible.

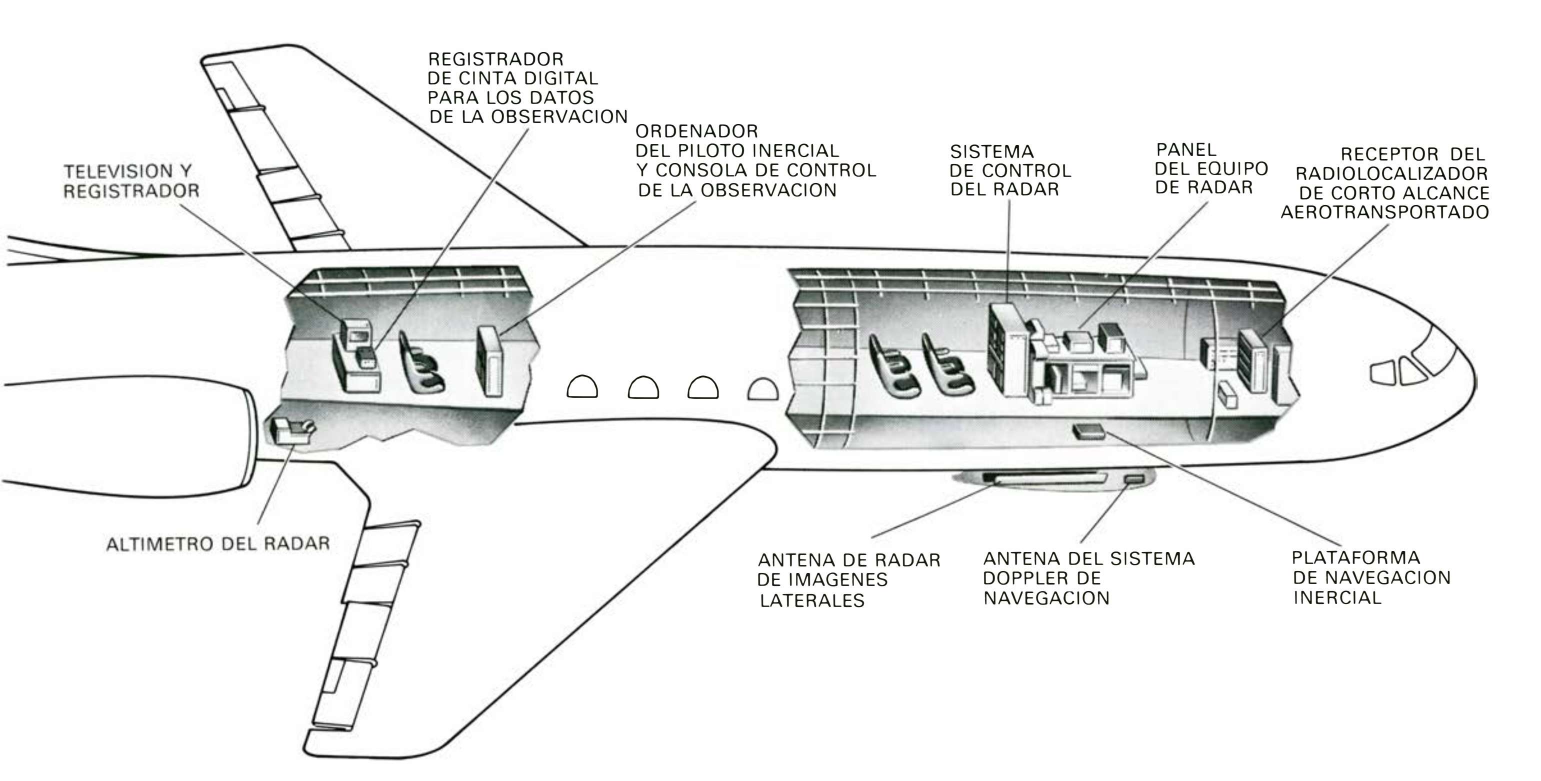
Los reflectores especulares por excelencia son las superficies lisas de agua. En general, estas superficies no son perpendiculares al haz del radar y, en consecuencia, reflejan toda la energía de las microondas que reciben hacia el espacio y ninguna fracción de esta energía retorna a la antena del radar. Como resultado de este hecho, las imágenes de los ríos y lagos obtenidas mediante un sistema de radar lateral suelen ser totalmente negras. Por el contrario, las superficies horizontales y verticales relacionadas entre sí (tales como las de un edificio junto a una carretera) pueden actuar en forma de diedro reflector, devolviendo a la antena una fracción importante de la energía emitida. Por consiguiente, estas superficies pueden parecer en el receptor



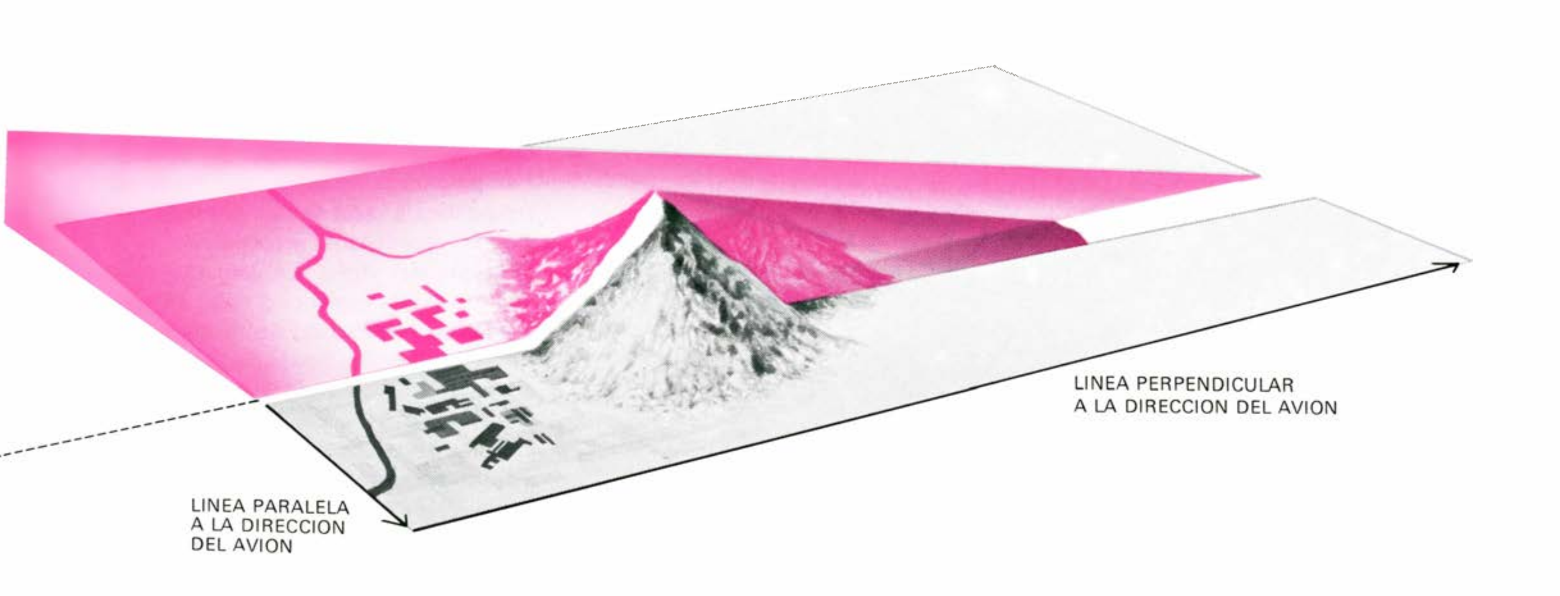
EL SISTEMA DE RADAR de imágenes laterales construido por la Goodyear Aerospace Corporation está instalado a bordo de un avión Caravelle, propiedad de la Aero Service Divi-



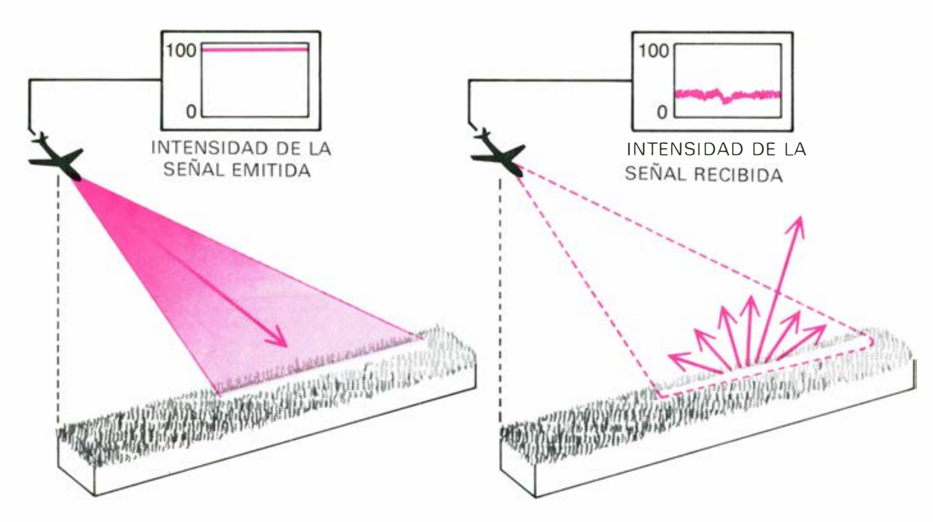
GEOMETRIA de un radar de imágenes laterales para el caso en que el avión sobrevuele un terreno montañoso. La antena fijada en el fuselaje emite impulsos cortos de microondas. El haz forma un abanico de pequeña apertura y su direc-



sión, de la Western Geophysical Company of America. El vuelo del avión está controlado con gran precisión por un sistema de navegación inercial especialmente programado. Este dispositivo controla también muchos aspectos del sistema de radar. Otras cámaras situadas a bordo del aparato (cámaras de televisión, cartográficas y multiespectrales) proporcionan una información adicional complementaria sobre el terreno observado, tanto para longitudes de onda visibles como para infrarrojas. Esta ilustración, al igual que la siguiente, es obra de George V. Kelvin.



ción es perpendicular a la dirección del vuelo. Cuando uno de dichos impulsos llega a tierra, es reflejado por ella y una fracción del impulso original vuelve al avión. Allí es recibido por la misma antena que lo emitió y es enviado a un receptor de radio muy sensible. En el receptor se genera una señal cuya amplitud depende de la intensidad de las microondas recibidas en cada instante. Esta señal controla la luminosidad del haz de un tubo de rayos catódicos. El haz se registra sobre una película. Puesto que las microondas se mueven en línea recta, las zonas tales como la ladera de la montaña más alejada no reciben y no devuelven los impulsos, y, por tanto, en las imágenes del radar aparecen en negro.



UN REFLECTOR DIFUSO, un campo de maíz, por ejemplo, difunde la energía de los impulsos de microondas en todas direcciones, devolviendo una fracción de la energía a la antena del radar.

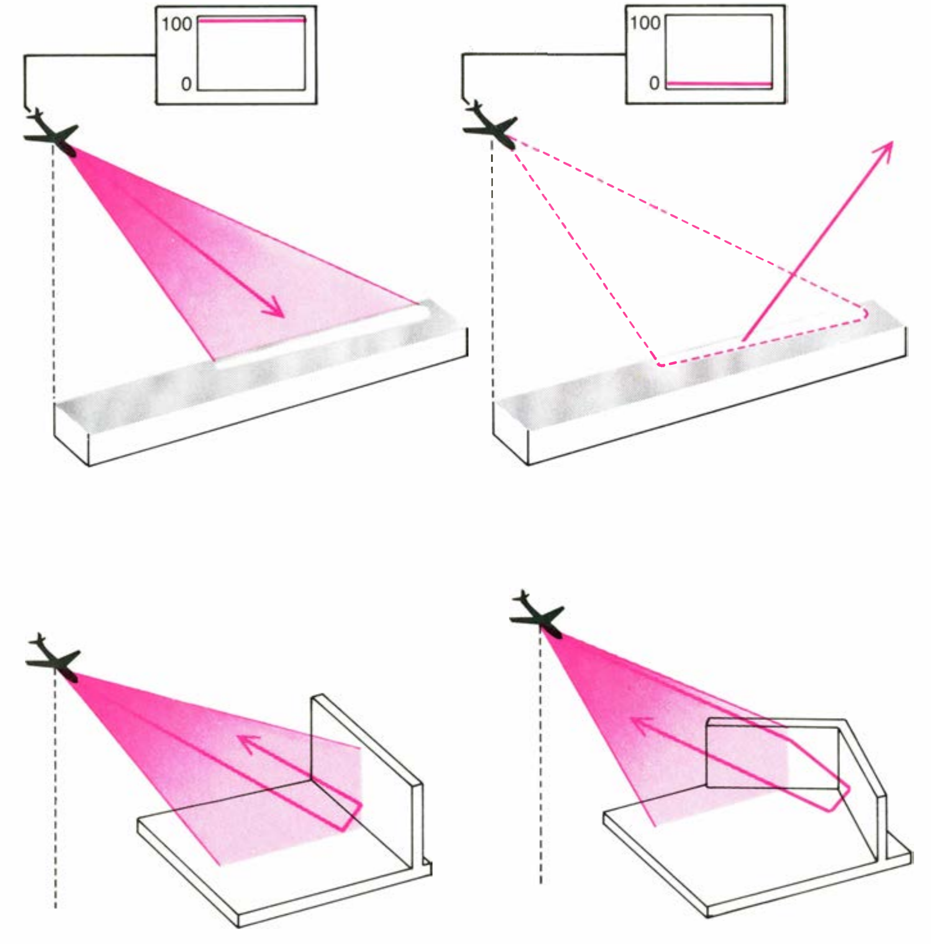
del radar miles de veces más brillantes que las superficies difusas de un tamaño semejante. En las zonas urbanas existen muchos de estos reflectores múltiples para las longitudes de onda situadas en la región de las microondas.

La geometría de una imagen fotográfica viene determinada por el hecho de que la relación angular de los rayos que llegan a una lente procedentes de un determinado objeto o escena se mantiene por parte de los que salen de ella y, finalmente, impresionan la película, como si todos hubiesen pasado a través de un diminuto agujero común. El tamaño de la imagen de un objeto distante es inversamente proporcional a la distancia entre el objeto y la lente. Las relaciones geométricas existentes entre los tamaños de los diferentes objetos situados en un mismo plano perpendicular al eje de la lente y colocado delante de ella se mantienen sobre un plano imagen, paralelo al anterior y situado al otro lado de la lente; las relaciones existentes entre las superficies inclinadas respecto del plano objeto se reproducen mediante relaciones trigonométricas sencillas sobre el plano imagen.

Un sistema de radar lateral aerotransportado funciona de acuerdo con unas consideraciones geométricas diferentes. Una de las coordenadas de la imagen que proporciona este sistema es paralela a la dirección en que avanza el avión; la otra coordenada es perpendicular a dicha dirección. El proceso de observación se basa en la emisión de un impulso de microondas de corta duración procedente de un transmisor de alta potencia. El correspondiente haz, en forma de abanico vertical, es emitido por una antena de gran longitud situada en el fuselaje del avión y dirigida perpendicularmente a la dirección del vuelo. El impulso se mueve con la velocidad de la luz; el tiempo necesario para que este impulso alcance un objeto determinado y, tras ser reflejado, vuelva al avión es proporcional a la distancia entre éste y el objeto.

La intensidad de la señal reflejada controla la intensidad de un punto brillante que se mueve sobre la pantalla de un tubo de rayos catódicos con una velocidad sincronizada proporcionalmente. Las posiciones sucesivas del punto corresponden a otras tantas distancias según la coordenada perpendicular a la dirección del vuelo. En conjunto, cada impulso del transmisor del radar da lugar a una línea de barrido de intensidad variable a través de la pantalla del tubo de rayos catódicos. Las líneas de barrido se van registrando sucesivamente sobre una película fotográfica que se mueve perpendicularmente a la dirección de dichas líneas con una velocidad proporcional a la velocidad del avión. Como resultado, sobre la película se obtiene una imagen continua del terreno observado línea a línea.

Los diferentes detalles del relieve del terreno aparecen más o menos desplazados en la imagen a lo largo de la dirección perpendicular a la del vuelo del avión respecto a las posiciones que ocuparían si fueran observados verticalmente. La razón de este hecho estriba en que la fracción del impulso de micro-



UN REFLECTOR ESPECULAR, verbigracia, una superficie lisa de agua o una superficie pulimentada, se comporta como un espejo y refleja la energía de las microondas en una dirección bien definida. Si la superficie no es perpendicular a la antena, la energía no regresa a esta *(arriba)* y en la imagen del radar la superficie aparece negra. Las superficies que se hallan relacionadas entre sí, como las casas y las calles, actúan conjuntamente como reflectores angulares *(abajo)*.

ondas reflejada por las zonas más elevadas del terreno (y, por lo tanto, más próximas al avión) retorna a la antena antes que la energía reflejada por las zonas menos elevadas del terreno. En consecuencia, las imágenes de las zonas más elevadas del terreno, más próximas al avión, aparecen un tanto desplazadas hacia la línea de vuelo del aparato. Este desplazamiento por elevación es una de las características de las observaciones con radar de terrenos accidentados. En cierto modo, este efecto es análogo a la distorsión de la perspectiva que se pone de manifiesto en la fotografía ordinaria.

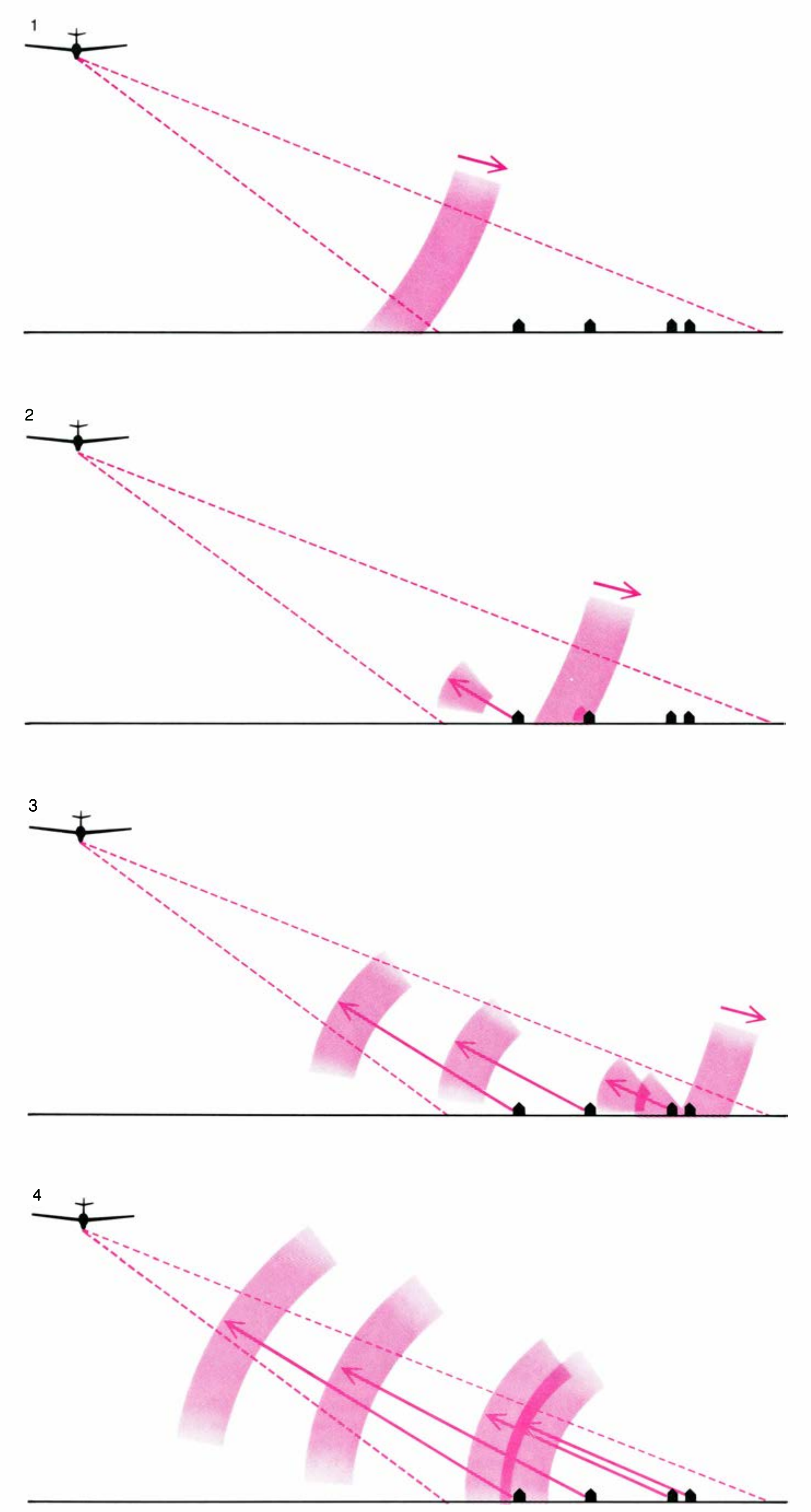
En este caso, todos los objetos con las mismas coordenadas angulares respecto a la lente parecen coincidir en la película. En un sistema de radar, todos los objetos situados a la misma distancia de la antena coinciden en la imagen. Por consiguiente, una imagen de radar es aproximadamente análoga a la imagen fotográfica que proporcionaría una cámara hipotética enfocada en una dirección perpendicular a la línea que une el objeto y la antena de radar.

La velocidad con que se mueve el avión es muy pequeña en comparación con la velocidad del impulso de microondas, y el haz en forma de abanico de la correspondiente radiación es siempre perpendicular a la línea de vuelo. Como consecuencia de este hecho, las imágenes en la dirección perpendicular a la línea de vuelo presentan una perspectiva normal (los objetos aparecen desplazados hacia el avión en función de su altura); en cambio, las imágenes en la dirección paralela a la línea de vuelo carecen totalmente de perspectiva. Así, mientras la escala y la apariencia de la imagen del radar en la dirección perpendicular a la línea de vuelo vienen determinadas por la velocidad de la luz, la escala y la apariencia de dicha imagen en la dirección paralela al vuelo vienen determinadas por la velocidad del avión. La operación de reconciliar e igualar estas dos escalas independientes sobre la imagen final es una de las fases más importantes de la obtención de mapas de la superficie terrestre mediante la utilización de sistemas de radar.

En la práctica, este problema y otros muchos relacionados con la obtención de imágenes de radar satisfactorias se ha resuelto instalando en el avión un piloto inercial. Este dispositivo utiliza un computador muy versátil, programado de forma tal que sea capaz de proporcionar toda la información necesaria para que el radar funcione correctamente. El piloto inercial controla el vuelo del avión, manteniendo una altura operativa adecuada



IMAGEN DE RADAR DEL RIO OHIO, al este de Ohio y Kentucky. Muestra el diferente comportamiento de las reflexiones difusas y especulares. La ciudad de Portsmouth está representada por el conjunto de reflectores especulares que aparecen sobre la orilla norte del río en el ángulo superior izquierdo. Al oeste de Portsmouth, el río Scioto se une con el río Ohio. Obsérvese que el agua de ambos ríos aparece negra. La carretera que discurre a lo largo del pie de las colinas, al oeste del río en Kentucky, es la autopista número 23. La tierra al este del río corresponde al Wayne National Forest. Esta imagen de radar pone de manifiesto la naturaleza montañosa del terreno y delimita muy claramente el curso del río. Los dos puentes que cruzan el río por el norte se comportan como reflectores especulares de las ondas de radar, al igual que las esclusas y la presa en el centro de la figura. Las manchas blancas alargadas que se ven en el centro del río son gabarras. Los cortes rectos a través de las colinas cubiertas de árboles son zonas despejadas para el paso de líneas de conducción eléctrica. La superficie representada en esta imagen es de 16 por 24 kilómetros. Esta figura es una pequeña parte de un mosaico de imágenes que cubre una superficie de unos 100.000 kilómetros cuadrados de Ohio, Kentucky y Virginia. El reconocimiento se llevó a cabo con el fin de estudiar las características del terreno asociadas con las reservas de gas natural.



LA DISTANCIA ENTRE LOS OBJETOS Y EL AVION viene determinada por el tiempo necesario para que los impulsos de microondas, cuya velocidad es igual a la de la luz, lleguen al objeto y vuelvan a la antena. La figura pone de manifiesto el hecho de que los impulsos reflejados por las dos casas más próximas a la antena llegarán a ésta antes que los procedentes de las dos casas más alejadas *(1 y 2)*. La resolución en una dirección perpendicular a la del vuelo del avión resulta igual a la mitad de la duración de cada impulso. Dos objetos, tales como la tercera y cuarta casas, que se hallan separados en dicha dirección por una distancia menor que la mitad de la duración del impulso del radar, quedarán detectados por la antena como si se tratara de un objeto único *(3 y 4)*.

a lo largo de una serie de círculos de grandes dimensiones (tanto como 1000 kilómetros), sobre la base de que cada una de las correspondientes rutas sea prácticamente paralela a la anterior; la unión de las sucesivas bandas de imágenes permite obtener un mapa completo del terreno en forma de mosaico. El piloto inercial es capaz de guiar al avión sobre las nubes o a través de ellas, durante el día o la noche, sin necesidad de introducir ninguna corrección de origen visual o de radionavegación.

El piloto inercial mantiene el vuelo regular del avión y evita las maniobras erráticas, cuya ocurrencia daría lugar a imágenes distorsionadas. Estos cometidos se llevan a cabo mediante el envío de las oportunas señales desde este dispositivo hasta el piloto automático que gobierna el avión. Por otra parte, los sensores angulares del piloto inercial controlan la posición de la antena de radar respecto a tres coordenadas angulares, con el fin de garantizar que la antena esté dirigida constantemente hacia el suelo formando un cierto ángulo con la línea de vuelo del aparato. El sistema inercial determina tanto el ritmo de emisión de los impulsos de radar como la velocidad con que avanza la película en que se registran las oportunas imágenes. La velocidad con que el punto luminoso barre la pantalla del tubo de rayos catódicos controla la escala perpendicular a la línea de vuelo y la velocidad de la película en que se registran las imágenes controla la escala paralela a la línea de vuelo.

Apertura sintética

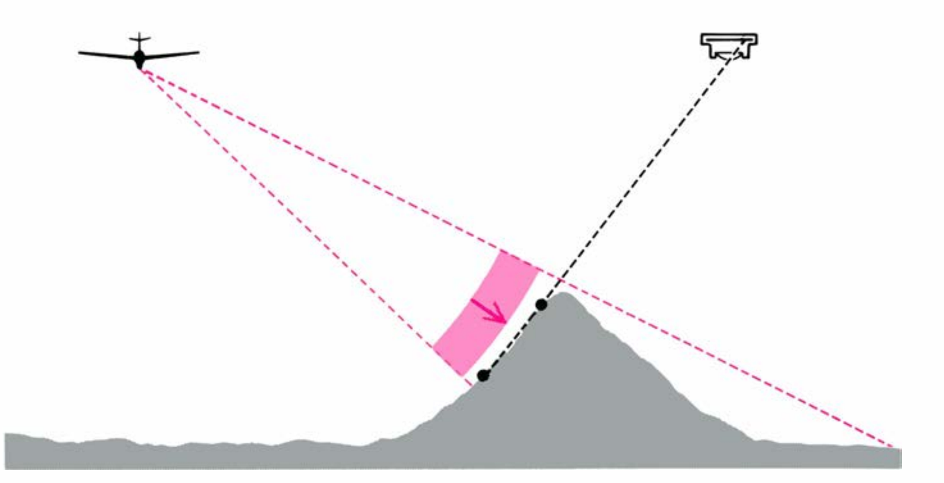
Dado que un sistema de radar lateral aerotransportado mide la distancia a que se encuentran los objetos del terreno por medio de impulsos de microondas, la resolución de la imagen en la dirección perpendicular a la línea de vuelo es proporcional a la duración de dichos impulsos. Las señales reflejadas por dos objetos diferentes se pueden resolver si la distancia que los separa es mayor que la semilongitud del impulso en cuestión. Por ejemplo, un impulso cuya duración sea de 10^{-7} segundos tiene una longitud de 30 metros; el poder de resolución de este impulso es igual a 15 metros. Para aumentar la resolución, es necesario reducir la duración del impulso.

Sin embargo, la resolución de la imagen en la dirección paralela a la línea de vuelo es proporcional a la anchura del haz de microondas que emerge de la antena. Dos objetos situados a la misma distancia de la antena y separados entre sí en la dirección paralela a la línea de

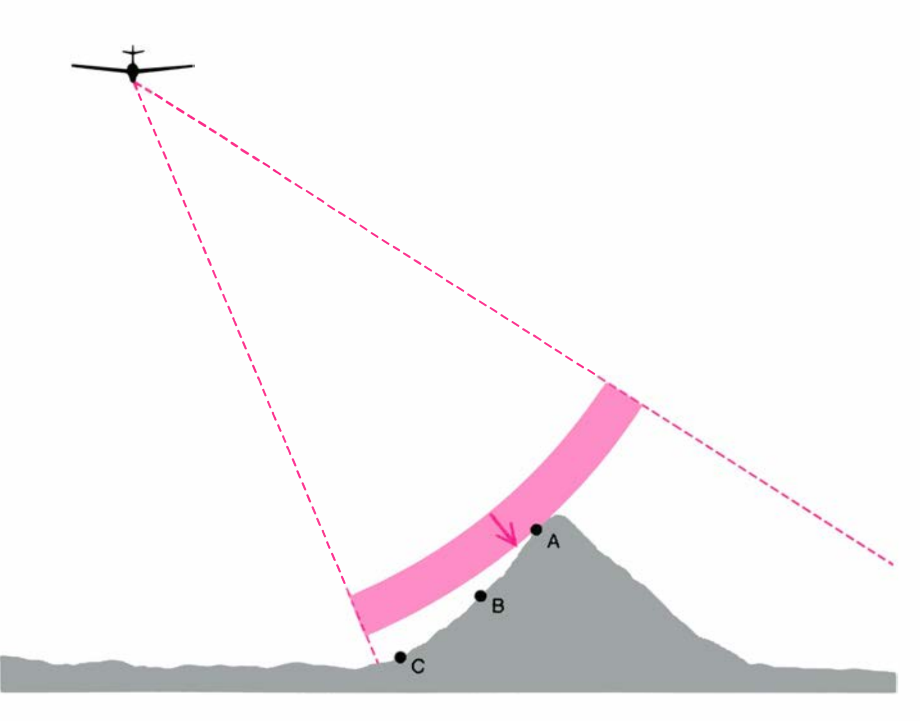
vuelo, se podrán resolver sólo en el caso de que su separación mutua sea mayor que la anchura del haz de radar para la distancia a la que se encuentran. En caso contrario, los dos objetos serán alcanzados simultáneamente por el citado haz, los impulsos de microondas reflejados por ellos llegarán a la antena al mismo tiempo y, finalmente, ambos aparecerán sobre la pantalla en forma de una imagen única e intensa. En consecuencia, la anchura angular del haz emitido por la antena determina la resolución en la dirección en que avanza el avión. Puesto que dicho ángulo se mantiene esencialmente constante al variar la distancia antena-objeto, la anchura del haz aumenta a medida que lo hace dicha distancia y, por lo tanto, la resolución en la dirección paralela a la línea de vuelo empeora.

La anchura angular del haz emergente es inversamente proporcional a la longitud de la antena. En consecuencia, la resolución aumenta considerablemente si utilizamos una antena de grandes dimensiones con una anchura del haz pequeña. Puesto que la longitud física de la antena que puede montarse sobre un avión resulta limitada por el tamaño del avión, aun en el caso de utilizar una antena de la máxima longitud posible en la práctica, la resolución de un sistema de radar lateral convencional es mucho peor en la dirección de la línea de vuelo que en la dirección perpendicular a ella, excepto para alcances muy pequeños.

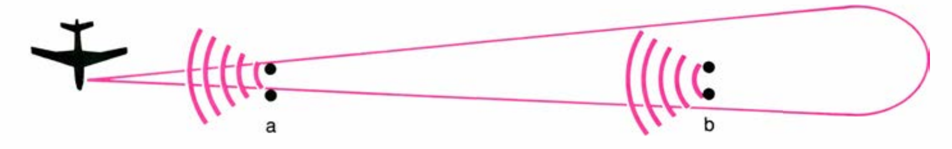
Sin embargo, aprovechando el movimiento del avión, se puede conseguir que una antena relativamente pequeña y de haz ancho "sintetice", o se comporte como una antena muy larga con un haz de poca anchura. Este efecto se puede lograr de la forma siguiente. La pequeña antena real transmite impulsos a intervalos regulares a lo largo de la dirección del vuelo. Cuando el avión sobrevuela un objeto determinado del terreno, nos encontramos con que el objeto entrá en el haz emitido, se mueve a través de él y, finalmente, lo abandona; durante todo este tiempo de tránsito, el objeto en cuestión está sometido al haz y refleja una serie de impulsos de microondas. Ahora bien, cuanto mayor es su distancia de la antena, tanto más tiempo permanece dentro del haz emitido por ella. En otras palabras, desde el punto de vista del terreno, la antena parecerá más o menos grande, respectivamente, según los objetos observados estén más o menos alejados de ella. De hecho, la longitud efectiva de la antena es directamente proporcional a la distancia que la separa del objeto considerado. Puesto que la resolución es directamente proporcional a la



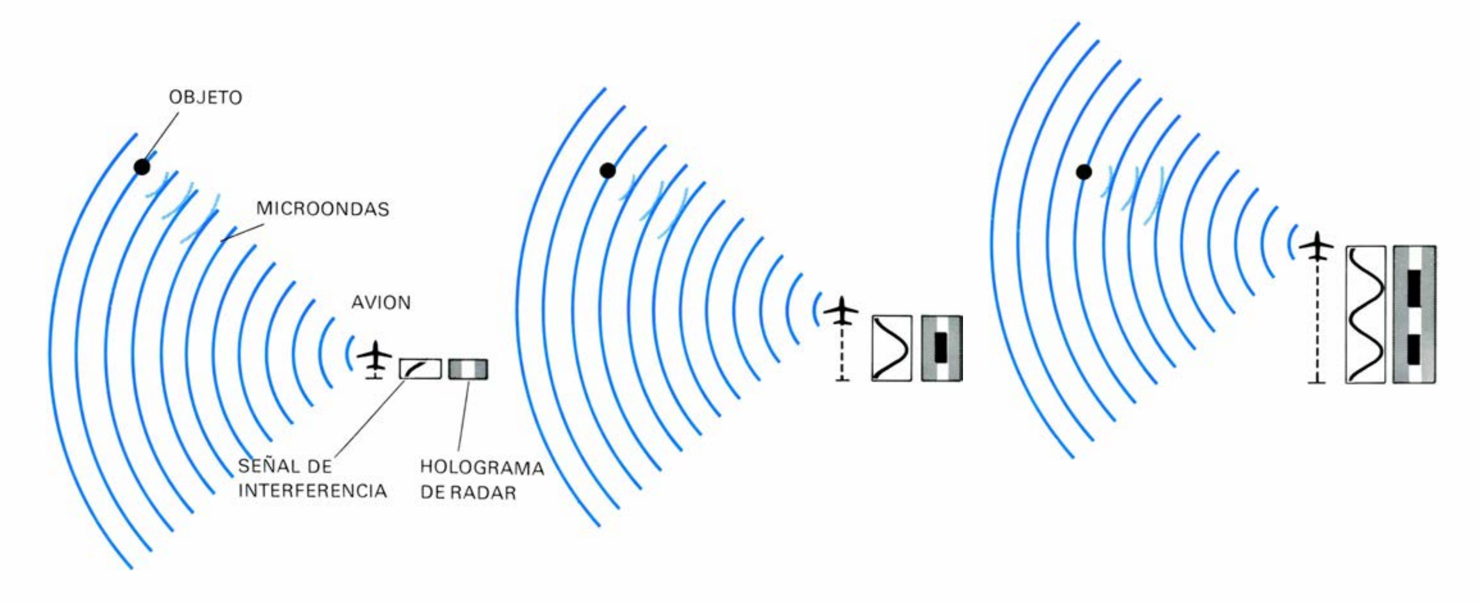
LA PERSPECTIVA EN LAS IMAGENES DE RADAR es un tanto diferente de la perspectiva en las fotografías aéreas ordinarias realizadas con luz visible. En las imágenes de radar, dos objetos cuya separación mutua sea menor que la anchura del correspondiente haz y cuya distancia a la antena sea la misma, aparecerán juntos. La razón de este hecho estriba en que los impulsos de microondas llegarán a ellos al mismo tiempo y volverán a la antena también al mismo tiempo, con lo cual los dos objetos se verán como uno solo. En fotografía, dos objetos parecen coincidir si sus coordenadas angulares respecto a la lente coinciden. En consecuencia, la perspectiva de las imágenes de radar es aproximadamente análoga a la de una cámara, situada en el mismo plano vertical en que se hallan el avión y los objetos, según una dirección perpendicular a la del haz del radar.



LOS OBJETOS ELEVADOS RESULTAN DESPLAZADOS hacia la línea de vuelo del avión, dando lugar a que en las imágenes de radar las montañas parezcan más inclinadas de lo que lo son en realidad. En efecto, el pico de una montaña *(A)* está más próximo al avión que su ladera *(B)* o su base *(C)*; así pues, los impulsos procedentes del pico vuelven a la antena antes que los otros.



LA RESOLUCION EMPEORA CON LA DISTANCIA antena-objeto debido a que el haz de un sistema de radar de imágenes laterales de apertura real se emite en una dirección perpendicular a la del vuelo del avión y es más ancho para distancias grandes que para distancias pequeñas. Dos objetos próximos se verán separados si su distancia al avión es pequeña *(a)*, pero se verán como un solo objeto si dicha distancia es grande *(b)*. La resolución de una antena grande sería mejor, pero evidentemente existe un límite físico en la longitud máxima de la antena que se puede instalar en un avión.



SINTESIS DE UNA ANTENA LARGA a partir de una antena corta, aprovechando el movimiento del avión. Cuando un avión vuela en línea recta, una antena corta real montada en su fuselaje envía una serie de impulsos a intervalos regulares. Cada impulso consta de un tren de microondas coherentes *(curvas de color oscuro).* **Aunque la duración de los impulsos determina la resolución en la dirección perpendicular a la del vuelo del avión, es la longitud de onda de la radiación de microondas la que determina la resolución en la dirección paralela a la del vuelo del avión. Cuando un objeto** *(punto negro)* **entra en el haz emitido por la antena** *(izquierda),* **refleja una fracción de los impulsos que recibe. En determinados puntos de la trayectoria del avión, la distancia entre el objeto y la antena será igual a un múltiplo entero de la longitud de onda de las microondas; entre dos de dichos puntos no lo será. En la presente figura, el objeto está separado de la antena primero por 11 longitudes de onda** *(1),* **luego por 10** *(2),* **a continuación por 9** *(3)* **y, finalmente, por 8 $^1/_2$** *(4),* **en cuyo caso el objeto es perpendicular a la antena. A partir de aquí, la distancia entre el aeroplano y el objeto aumenta** *(5).* **La antena recibe la serie de ondas reflejadas** *(curvas de color claro)* **y las combina electrónicamente con un tren de longitudes de onda de referencia (no representado en la figura), lo que da lugar a que las dos series de oscilaciones interfieran. La señal de interferencia emerge en forma de tensión; esta tensión controla la luminosidad de un punto que**

longitud de la antena e inversamente proporcional a dicha distancia, en un radar de apertura sintética los dos efectos se compensan entre sí; la resolución de la imagen en la dirección paralela a la línea de vuelo se mantiene constante y, por lo tanto, es independiente de la distancia antena-objeto. El radar de apertura sintética permite obtener imágenes de alta resolución del terreno situado a muchos kilómetros de distancia del avión.

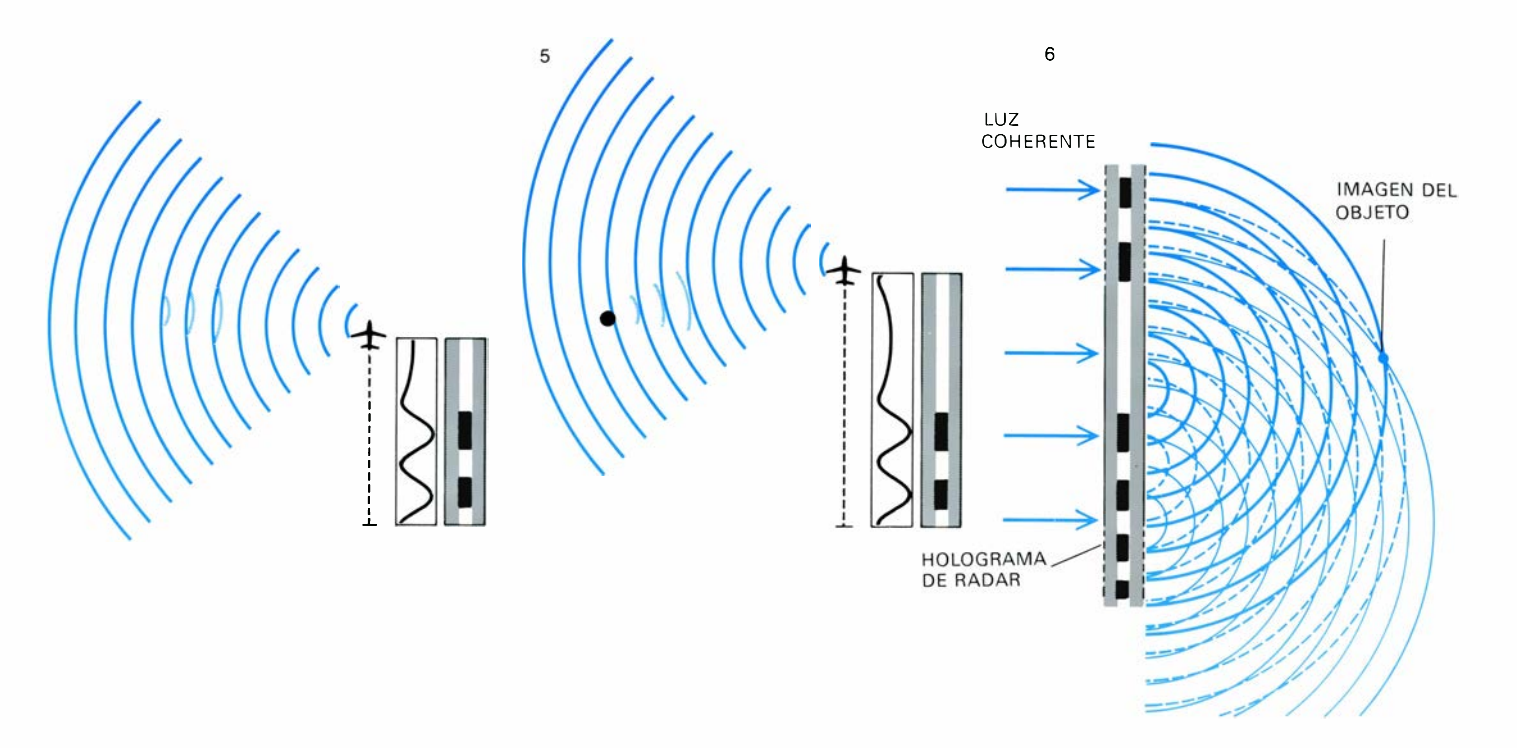
Holografía de radar

Entre un sistema de radar de apertura sintética y un sistema holográfico de luz visible existe una analogía muy interesante. La holografía consiste en un sistema de producción de imágenes en el que las ondas luminosas procedentes de una fuente de luz coherente tal como el láser iluminan un objeto o una escena. Cada punto del objeto refleja la luz y las ondas reflejadas se hacen interferir con un haz de ondas luminosas de referencia. Las pautas circulares de interferencia resultantes se registran en una película fotográfica de grano fino. Cuando la película revelada se ilumina con un haz de luz coherente, las pautas de interferencia reconstruyen el conjunto original de ondas reflejadas. Este conjunto constituye una imagen tridimensional del objeto o escena original. Esta imagen está caracterizada por un extraordinario detalle y fidelidad.

El radar de apertura sintética es un sistema de producción de imágenes en el que un haz de microondas coherente ilumina el terreno situado debajo y a uno de los lados del avión. El receptor de radar detecta las ondas difundidas por cada punto del terreno. En el equipo electrónico que transporta el avión, las señales generadas en el receptor por el tren de impulsos de microondas reflejados se combinan con un tren de señales de referencia con el fin de originar pautas de interferencia. Cada una de estas pautas se representa en la pantalla de un tubo de rayos catódicos, línea a línea, y se registra sobre una película móvil cuya velocidad es proporcional a la velocidad del avión. Esta película de datos puede ser considerada como un holograma de radar.

En un radar de apertura sintética, la pauta de interferencia producida por un difusor determinado del terreno no queda registrada en forma de una figura circular (como sucede en la holografía), sino en forma de una línea estrecha y discontinua paralela al borde de la película. Ahora bien, se puede considerar que dicha línea representa la sección transversal de una pauta de interferencia circular; la anchura de la línea está relacionada con la duración de los impulsos de radar. La distancia difusor-antena queda registrada en cada caso a través de la distancia que separa a la línea en cuestión del borde de la película de registro. Dos difusores situados a la misma distancia de la antena y lo suficientemente próximos entre sí como para resultar "iluminados" simultáneamente por el haz de radar, darán lugar a dos secuencias de líneas discontinuas, ligeramente desplazadas entre sí y un tanto solapadas la una con la otra. Igual que sucede en los hologramas de luz visible, las dos secuencias de figuras de interferencia se pueden superponer sin destruir su capacidad de dar lugar a imágenes independientes.

Cuando la película de grabación se ilumina con un haz de luz coherente, la pauta de interferencia originada por cada



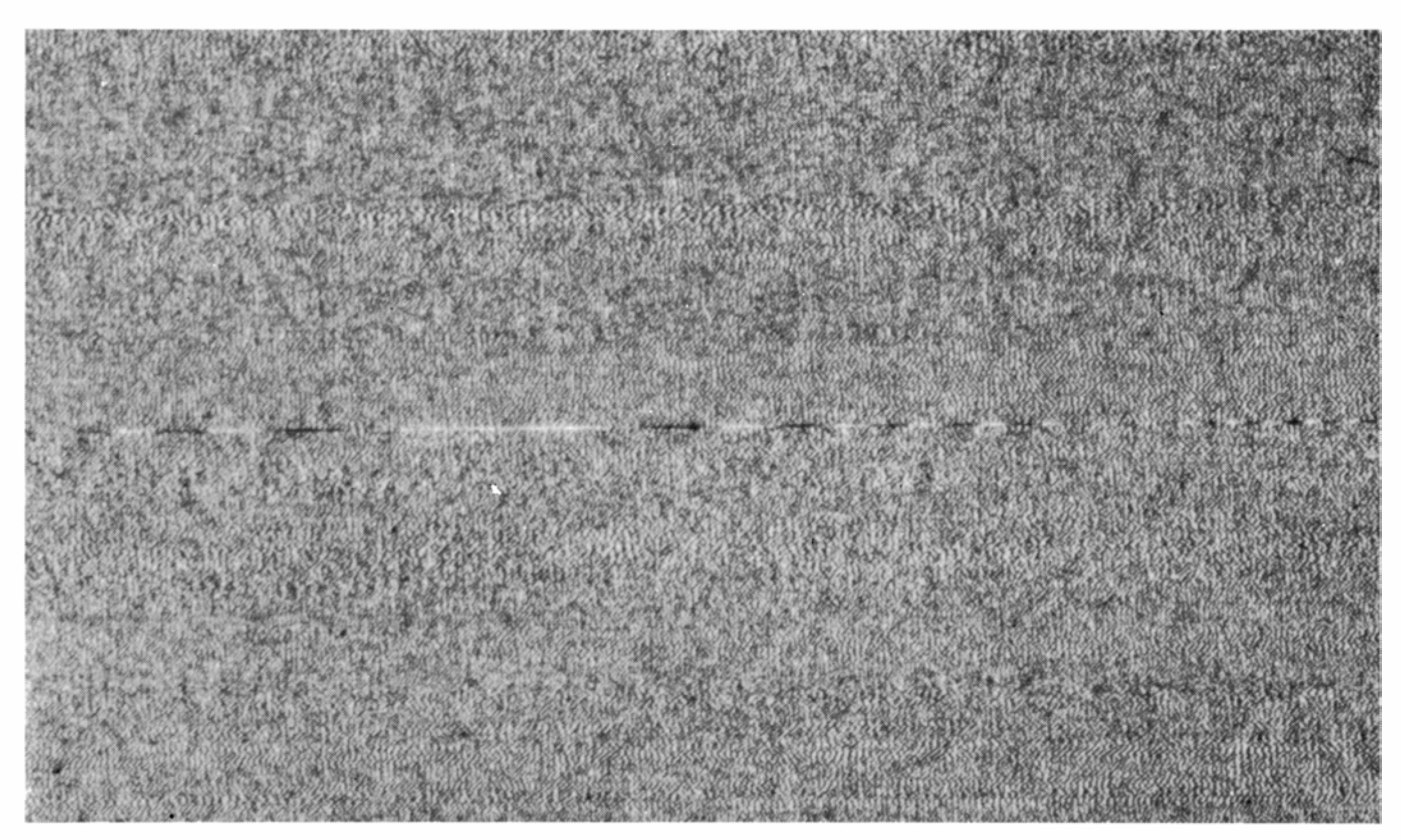
barre la pantalla de un tubo de rayos catódicos. Cuando el impulso reflejado coincide con el impulso de referencia, la interferencia es constructiva; la tensión será alta y el citado punto será luminoso. Cuando la fase de la onda reflejada no coincida con la de la onda de referencia, la interferencia será destructiva; la tensión será baja y el punto móvil será obscuro. En consecuencia, dicho punto traza una serie de rayas claras y obscuras de longitudes desiguales. Estas imágenes se registran sobre una película que se mueve con una velocidad proporcional a la del avión. De hecho, la serie de rayas opacas y transparentes de la película es una figura de interferencia monodimensional; la película en cuestión es un holograma de radar. Cuando este holograma se ilumina mediante una fuente de luz coherente *(6)*, **cada raya transparente se comporta como una fuente individual de luz coherente. Al otro lado del holograma existirá un punto único en el que todas las ondas luminosas resultantes interfieran constructivamente. En este punto, la undécima longitud de onda de la luz** *(curvas delgadas en color)* **procedente de la línea transparente originada por la undécima microonda, se encuentra con la décima longitud de onda de la luz** *(curvas discontinuas en color)* **procedente de la línea transparente originada por la décima microonda y, a su vez, ambas se encuentran con la novena longitud de onda de la luz** *(curvas gruesas en color)* **procedente de la línea transparente originada por la novena microonda. En dicho punto se forma una imagen en miniatura del objeto original.**

difusor convierte una pequeña fracción de la luz transmitida a través de la película en una onda divergente susceptible de ser focalizada mediante una lente adecuada. Existe la posibilidad de transformar muchas señales solapadas originadas por muchos difusores simultáneos en una serie de imágenes diferenciadas. De este modo, se forma una réplica en luz visible de las ondas de radar recibidas por la antena; en última instancia, lo que se obtiene es una imagen en miniatura del terreno observado. Puesto que en el holograma de radar están contenidos todos los datos procedentes de cada punto y de cada objeto del terreno para toda la longitud de la antena sintética, la resolución de la imagen holográfica corresponde a la resolución de la antena de pequeña apertura real. En consecuencia, la resolución de la imagen formada por el holograma es excelente.

Una forma adecuada de registrar la imagen del holograma consiste en hacer avanzar la película de grabación de datos a través de un haz de luz coherente, de forma tal que el movimiento de la película de registro de imágenes esté sincronizado con el de las imágenes. Se produce así un registro continuo del terreno a lo largo de toda la ruta del avión. Si algún lector está familiarizado con la holografía, podría suponer que el sistema óptico utilizado para obtener la película con la imagen final no presenta ninguna dificultad, puesto que la imagen del holograma se forma simplemente iluminando dicho holograma con un haz de luz coherente. Sin embargo, en el caso de un holograma de radar, el problema no es tan sencillo como parece a primera vista.

En la práctica, el registro de las imágenes en cuestión presenta dos dificultades. Una de ellas consiste en el hecho de que el aspecto holográfico del registro de los datos se pone de manifiesto solamente en la dirección longitudinal de la película. Las señales ya están resueltas de acuerdo con sus alcances respectivos a lo largo de la película. Como resultado del proceso de focalización de las señales en una dirección paralela a la del vuelo, las señales en la dirección perpendicular a la del vuelo se desfocalizan. Este problema se corrige introduciendo una lente cilíndrica entre la película de grabación de los datos y la película de registro de las imágenes; el poder de focalización de la lente cilíndrica se ejerce solamente en la dirección perpendicular a la del vuelo del avión. Mediante esta compensación óptica es posible obtener imágenes perfectamente enfocadas en ambas direcciones.

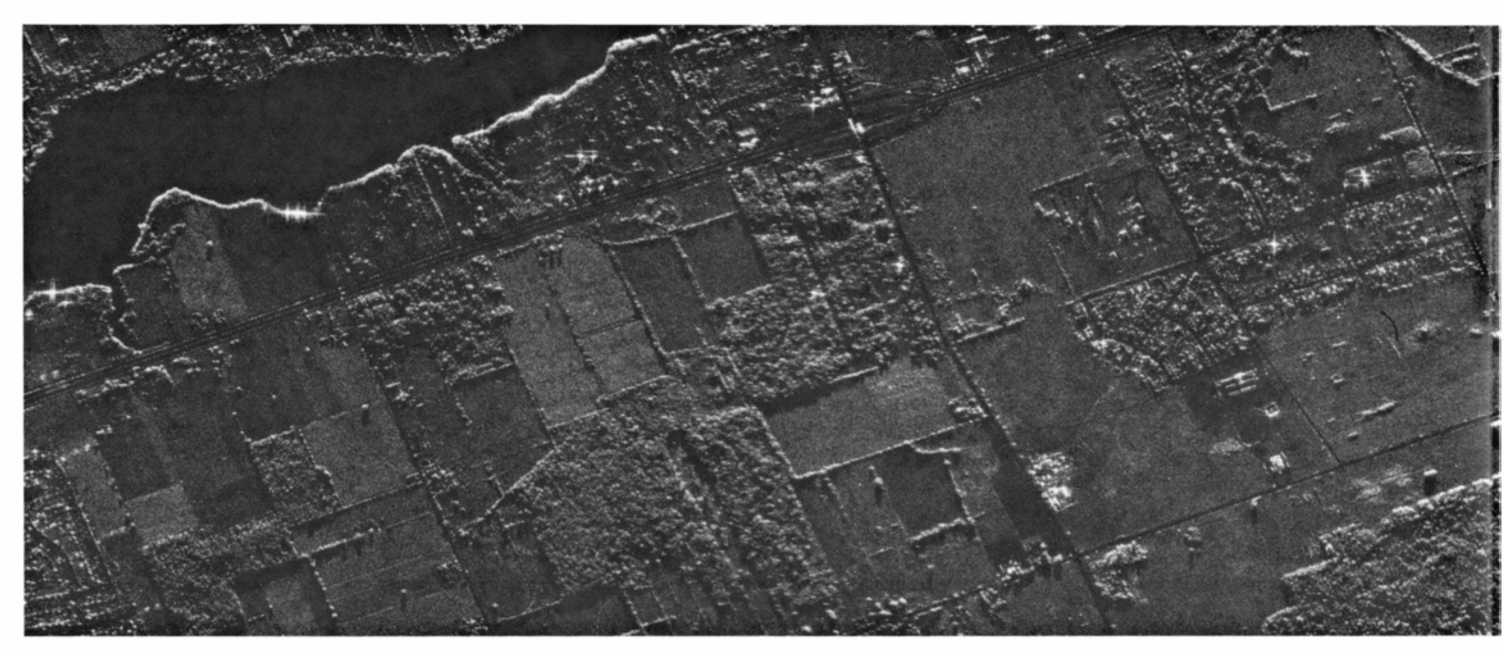
El otro problema consiste en que cuando se ilumina la película de grabación de datos, las señales correspondientes a los objetos muy alejados de la antena dan lugar a frentes de onda cuyos radios de curvatura son mayores que los de las señales de los objetos cuya distancia a la antena es pequeña. En consecuencia, las señales procedentes de los extremos opuestos de la película de grabación de datos focalizan la luz a diferentes distancias de la película. La distancia entre la película y el foco depende de lo alejado que se encuentre el objeto. La imagen obtenida de este modo está inclinada y si una cámara tratara de registrar esta imagen no sería posible enfocarla simultáneamente en su totalidad. Realmente, no podemos decir que este problema sea inesperado; después de todo, el terreno sometido a observación se encuentra en una posición extremadamente oblicua respecto a la antena de radar del avión



FRAGMENTO DE UNA PELICULA de registro de datos que muestra las figuras de interferencia producidas por los trenes de microondas reflejados en el radar de imágenes laterales de apertura sintética. El holograma del radar ha sido ampliado 10 veces para hacer visibles los detalles. La larga línea horizontal de trazos blancos y negros que aparece en el centro de la ilustración es la figura de interferencia de una onda intensa procedente de un reflector especular particularmente grande, verbigracia, una de las torres de conducción eléctrica de la imagen de radar que se reproduce en la portada. Las variaciones menores y menos organizadas corresponden a otras ondas reflejadas, menos intensas, procedentes de objetos menores.

y, por lo tanto, la imagen resultante estará también inclinada. El problema se puede corregir colocando una lente cónica frente a la película de grabación de datos. Los sistemas ópticos utilizados en la reconstrucción de los hologramas de radar y las lentes de corrección empleadas en la toma de imágenes se cuentan entre los más raros proyectados nunca; a pesar de ello, la utilización de estos sistemas y lentes no ofrece demasiadas dificultades en la práctica.

El primer proyecto de realización de un mapa de terreno a gran escala mediante un sistema de radar lateral aerotransportado consistió en un reconocimiento completo de la provincia de Darién, que enlaza Panamá con Sudamérica. En el año 1968, la Westinghouse Electric Corporation, en colaboración con la Raytheon Company, utilizó un sistema



FRAGMENTO DE UNA IMAGEN, registrada por un radar de imágenes laterales por el Environmental Research Institute de Michigan, que muestra la existencia de un claro en el terreno próximo a Detroit y Ann Arbor. El norte está en la parte inferior de la imagen. De acuerdo con esta orientación, el avión volaba según la dirección horizontal de esta figura. El agua del Lago Belleville y las pistas del aeropuerto de

de radar lateral de apertura real que la Westinghouse había desarrollado para el Ejército de los Estados Unidos, obteniendo un mosaico de imágenes que cubría una superficie de 20.000 km². El mapa completo de esta región no había sido obtenido con anterioridad a este trabajo debido a que la zona está cubierta casi siempre por una capa de nubes. La Westinghouse y, más tarde, la Grumman Ecosystems llevaron a cabo otros reconocimientos utilizando un sistema de radar desarrollado por la empresa Motorola Inc., también para el Ejército. El equipo de la Motorola, con algunos cambios, se utiliza actualmente en los reconocimientos que lleva a cabo la Motorola Aerial Remote Sensing Inc.

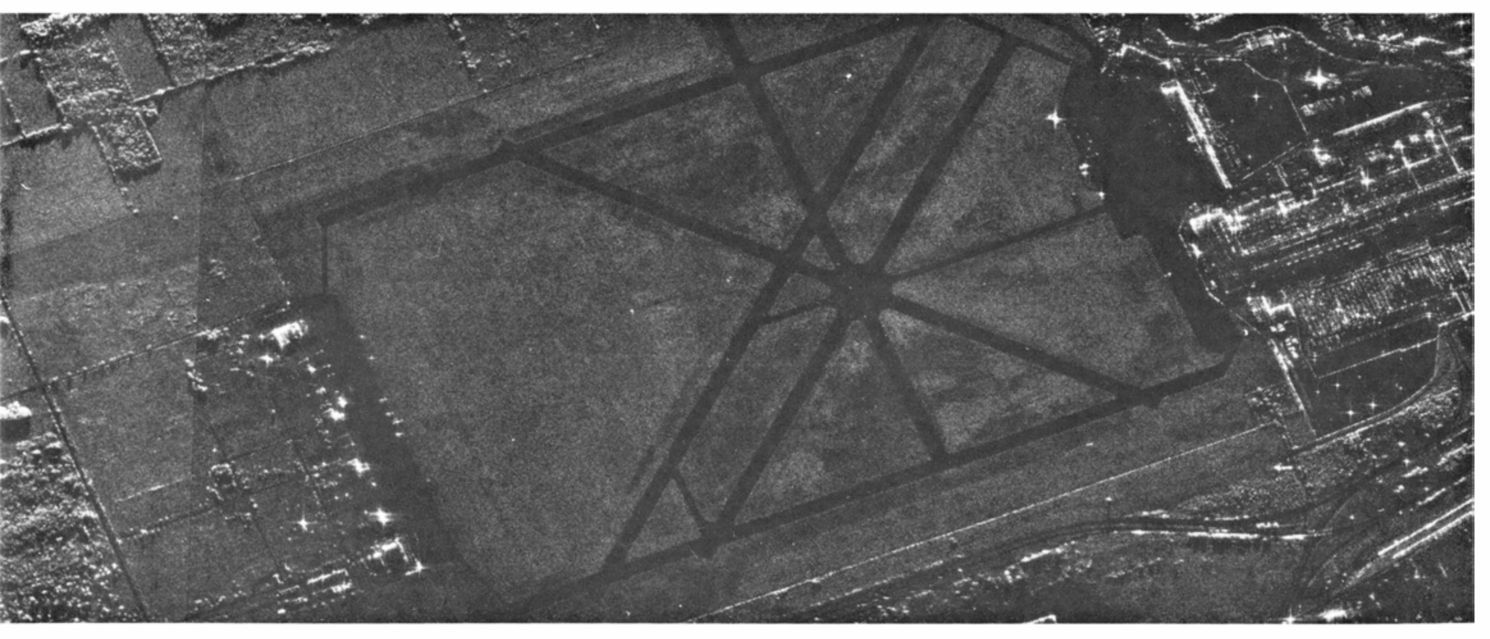
Aplicaciones del sistema

En 1970 los sistemas de radar de apertura sintética dejaron de estar sometidos a una clasificación de materia reservada o perteneciente a la seguridad militar. Un grupo de la Goodyear Aerospace Corporation y la Aero Service Division de la Western Geophysical Company of America, empresa subsidiaria de la Litton Industries, llevaron a cabo una serie de trabajos de reconocimiento de carácter civil utilizando un sistema de radar de apertura sintética construido por la Goodyear e instalado a bordo de un avión Caravelle perteneciente a la Aero Service. Otros sistemas de reconocimiento análogos a éste fueron construidos y explotados por un grupo de investigadores de la Universidad de Michigan. El Environmental Research Institute de Michigan se encarga actualmente de llevar a cabo los programas de este grupo. Otras varias organizaciones han contribuido también significativamente al desarrollo de este campo de trabajo.

En 1971 las grandes selvas de Sudamérica, principalmente en la región del Amazonas, eran las últimas regiones del mundo deficientemente conocidas desde el punto de vista cartográfico. En relación con los planes de desarrollo de sus regiones selváticas, Venezuela y Brasil se interesaron por el reconocimiento con radar de grandes superficies de sus territorios. Había que hacer frente a diversos requisitos. Las imágenes de todo el terreno observado tenían que ser coherentes entre sí, con el fin de que la comparación entre las diferentes regiones fuera fácil. Por otra parte, la información proporcionada por el sistema de radar debía ser coherente con la que proporcionaba el Earth Resources Technology Satellite (conocido actualmente con el nombre de Landsat) que en aquel entonces estaba llevando a cabo un reconocimiento general de la tierra. De acuerdo con estas consideraciones, el grupo Goodyear/Aero Service realizó todos sus vuelos de reconocimiento en la dirección norte-sur, con la antena enfocada hacia el oeste. De este modo, el aspecto de la tierra y las sombras sería muy parecido al de las imágenes del Landsat, que estaba preparado para fotografiar toda la tierra aproximadamente a las 9 de la mañana (hora local), cuando el suelo está iluminado oblicuamente por los rayos del sol.

Puesto que las imágenes del radar se iban a utilizar para confeccionar mapas, se procuró que la precisión de las posiciones fuera lo mayor posible desde un punto de vista práctico. El piloto inercial, totalmente adecuado para el control del vuelo del avión, no permitía determinar con la suficiente precisión los puntos de referencia de los mapas. En consecuencia, con ayuda del nuevo Transit Satellite Positioning System se fijaron las posiciones de unos 30 puntos de referencia distribuidos a lo largo de la región del Amazonas. La posición de estos puntos se midió con una precisión del orden de 10 metros. En cada uno de estos puntos se colocó un sistema de radiolocalización. El equipo existente a bordo del avión podía detectar las señales emitidas por este sistema hasta una distancia de 500 km. La recepción simultánea de las señales procedentes de dos de los dichos puntos permitía determinar la posición del avión con una precisión del orden de 75 metros; esta posición se comprobaba cada seis segundos. Durante el reconocimiento, la posición del avión se registraba en una cinta magnética, junto con su altura y otros datos significativos del vuelo. Con este sistema, en menos de un año se obtuvo un mosaico de imágenes de toda la Selva Amazónica del Brasil (una superficie de unos cuatro millones de kilómetros cuadrados, aproximadamente la mitad de la superficie de los Estados Unidos). La precisión de las imágenes de reconocimiento fue de unos 500 metros.

Con el fin de completar este trabajo de reconocimiento, Brasil puso en marcha el llamado Proyecto Radam (radar del Amazonas), con base en Belem. En



Willow Run aparecen en forma de manchas oscuras (en estos casos, el haz de radar se refleja especularmente). Los árboles y arbustos son reflectores difusos. Los edificios y las calzadas actúan conjuntamente dando lugar a reflectores angulares; la correspondiente reflexión objeto-antena es muy intensa en estos casos. La imagen aparece un tanto deformada debido a que la velocidad de la película de registro de datos no se midió bien.

el Proyecto Radam participaban geólogos, científicos especializados en el estudio del suelo y otros técnicos. Los mosaicos antes citados constituyeron el punto de partida de todos sus estudios. El correspondiente análisis geológico permitió obtener una serie de mapas detallados e hizo posible la realización de muchos trabajos de campo en ciertas zonas. Toda esta información permitió revisar y ampliar el conocimiento geológico de la región del Amazonas. Sobre la base de los mosaicos de imágenes obtenidos en estos trabajos se seleccionaron las alternativas posibles para la construcción de los diferentes tramos de la Autopista Transamazónica. Se descubrieron grandes depósitos de minerales importantes y se pusieron de manifiesto nuevos detalles geográficos desconocidos previamente, tales como conos volcánicos e incluso grandes ríos.

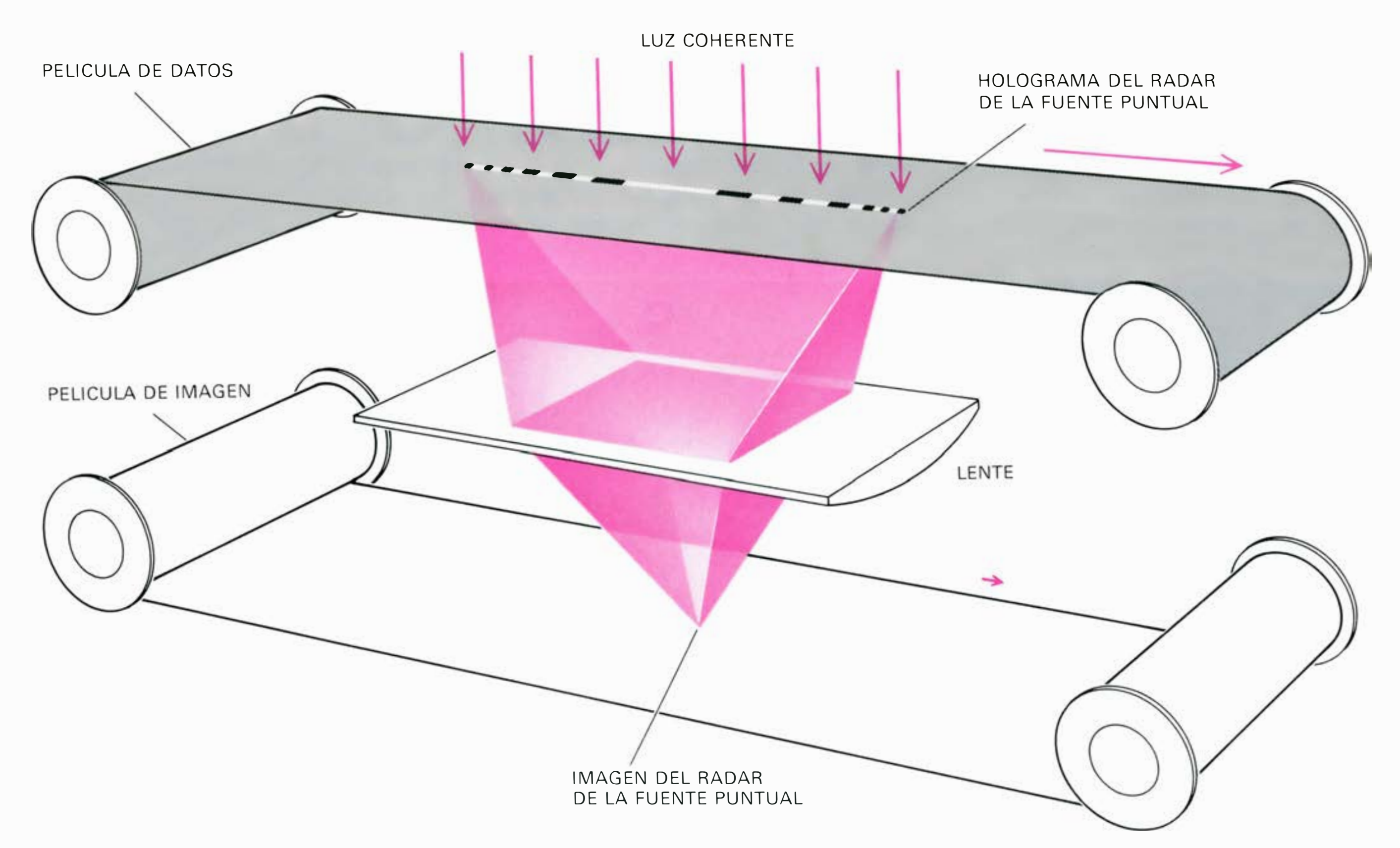
Hasta ahora, el Grupo Goodyear/Aero Service ha llevado a cabo trabajos de reconocimiento en las selvas del Amazonas del Brasil, Venezuela, Colombia oriental, Perú y Bolivia. Estos trabajos han acabado con las zonas blancas en los mapas de Sudamérica. Ningún otro sistema hubiera podido realizar este trabajo. A la vista del éxito alcanzado en estos reconocimientos, el trabajo se ha extendido a regiones de las que ya existían mapas obtenidos por otros medios. En marzo de 1976, la última serie de vuelos completó el reconocimiento con radar de todo el Brasil. Este trabajo cartográfico tiene una extensión sin precedentes.

Otras aplicaciones

El radar de imágenes laterales aerotransportado ha sido una herramienta de valor excepcional en otras aplicaciones. El estudio de los corrimientos continentales ha demostrado que es importante reconocer ciertos tipos de defectos, alineaciones y otros accidentes tectónicos. En muchos casos, estos accidentes son evidentes desde la superficie, pero son tan sutiles o se presentan en una escala tan grande que es muy difícil detectarlos. El pequeño valor del ángulo rasante del haz de microondas producido por un sistema de radar de imágenes laterales, que resalta las características morfológicas del terreno, y el enorme tamaño de las superficies que pueden ser reconocidas en condiciones constantes, favorecen la localización de los citados accidentes tectónicos. Dada la importancia que tiene la inexistencia de defectos geológicos activos en los lugares en que se sitúan las centrales nucleares, los sistemas de observación con radar pueden ayudar a seleccionar los emplazamientos adecuados. Por otro lado, debido a que las superficies lisas de agua producen un ennegrecimiento total en el radar, los ríos y zonas húmedas aparecen claramente representados en las correspondientes imágenes; en relación con los límites de una inundación o un desbordamiento sucede algo similar.

¿Cuáles podrían ser las aplicaciones del radar de imágenes laterales aerotransportado en el futuro? Entre las personas que trabajan en este campo existe la impresión general de que nuestra capacidad de producir imágenes de radar es mayor que nuestra capacidad de extraer información de ellas. En consecuencia, actualmente se está dedicando más atención a este segundo aspecto del problema.

Ciertas mejoras en las técnicas de exhibición de imágenes han aumentado considerablemente las posibilidades de extracción de información. Uno de los métodos consiste en utilizar un sistema de color que permita poner de manifiesto la enorme luminosidad del radar, cuyo intervalo energético supera



LA IMAGEN HOLOGRAFICA se restablece haciendo pasar la luz coherente de un láser a través de la película de registro de datos. A medida que la película avanza bajo dicho haz luminoso, la imagen reconstruida se registra sobre otra película en movimiento. Dado que la película de datos es holográfica únicamente según la coordenada paralela a la dirección de vuelo del avión, las imágenes según la coordenada perpendicular a dicha dirección deben focalizarse mediante una lente cilíndrica. La calidad de la imagen ha mejorado sensiblemente debido, fundamentalmente, a que la longitud de las figuras de interferencia sobre la película de datos es 72 veces mayor que la imagen final que dichas figuras dan.

con frecuencia las posibilidades de las películas en blanco y negro. Un segundo método utiliza técnicas estereográficas en relación con los datos representados en sendas imágenes de radar obtenidas de forma adecuada. Un tercer método consiste en capitalizar el hecho de que dos sensores remotos, operando conjuntamente, son generalmente mucho más efectivos de lo que sería uno solo. Los investigadores de la U.S. Geological Survey y la Goodyear Aerospace han trabajado juntos acerca de la posibilidad de combinar las imágenes que proporcionan los sistemas del radar lateral aerotransportado con las imágenes obtenidas por el Landsat; esta combinación se beneficia de las posibilidades de los dos sistemas. En esta combinación, las imágenes en color del Landsat se hacen más significativas en el contexto de la mejor presentación y mayor detalle que caracterizan a las imágenes de radar.

A un nivel más elaborado, se puede llegar a extraer una gran cantidad de información de los conjuntos multiespectrales de imágenes de radar obtenidos con longitudes de ondas diferentes y con señales de microondas polarizadas de distintas formas. Estudiando cómo varían las propiedades reflectoras de una superficie con la longitud de onda y la polarización se puede mejorar la percepción de diferencias entre las superficies observadas, y se pueden llegar a distinguir los diferentes tipos de suelos y los diferentes cultivos, por ejemplo.

El radar lateral puede contribuir también de forma significativa a la exploración de los planetas. Utilizando grandes sistemas de radar con base en la tierra ya se han obtenido fotografías de la superficie de la Luna y de Venus. La nave espacial Apolo 17 transportaba un sistema de radar de apertura sintética de tres longitudes de onda con el que tomó fotografías de la superficie de la Luna desde una altura de 100 km y con una resolución de 10 metros para la longitud de onda más corta utilizada (dos metros). Ahora bien, la superficie de la Luna nunca está cubierta de nubes y, en cambio, la superficie de Venus siempre lo está. Durante estos últimos años ha surgido la idea de situar un sistema de radar sobre un satélite de órbita alrededor de Venus y hacer que envíe las correspondientes fotografías a la tierra. Este sistema sería capaz de mostrar la superficie del planeta oculto con un gran detalle en zonas limitadas, complementando así la información que proporcionan las fotografías obtenidas con radares terrestres, caracterizados por una cobertura general y una resolución limitada.

Estructura y funcionamiento de los antígenos histocompatibles

Los antígenos de histocompatibilidad son proteínas de la superficie celular que causan el rechazo de los tejidos injertados de una persona en otra. También pueden defender al organismo contra infecciones y contra el cáncer

Bruce A. Cunningham

Cuando se transplanta un trozo de piel o un órgano de una persona a otra, el injerto no suele sobrevivir. Al cabo de una semana, aproximadamente, el tejido transplantado se inflama y poco después muere y se desprende. La medicina premoderna achacaba tales fracasos a la sepsis, pero actualmente se sabe que el sistema inmune de la persona que recibe el transplante reconoce como extrañas ciertas moléculas presentes en las células injertadas, y ataca y destruye dichas células del mismo modo que lo haría con bacterias u otros agentes invasores. Las moléculas que se reconocen como extrañas se denominan antígenos de histocompatibilidad o antígenos de transplante; proporcionan a los tejidos de cada individuo su propia identidad química.

Aunque, en un principio, el interés por estas moléculas de la superficie celular partió de su papel en el rechazo de injertos, ésta no puede ser su función biológica normal, ya que en la naturaleza no se da el trasplante de tejidos entre individuos. No obstante, la capacidad para rechazar injertos extraños puede remontarse a lo largo del árbol evolutivo hasta los vertebrados más primitivos, lo cual sugiere que los antígenos de histocompatibilidad confieren alguna ventaja importante para la supervivencia del huésped. Recientes experimentos han aportado pruebas de que estos antígenos se integran en un sistema de defensa inmunológico que protege al cuerpo de sus propias células cuando éstas se hallan infectadas por virus o se tornan cancerosas.

Los antígenos de histocompatibilidad se definen como moléculas que difieren de unos individuos a otros y se reconocen en el rechazo de injertos. En el ratón, se han detectado aproximadamente treinta antígenos de histocompatibilidad, cada uno de los cuales puede variar en poblaciones no isogénicas. Esta diversidad genética proporciona singularidad para cada individuo en una población normal. Los estudios de antígenos de histocompatibilidad ponen de manifiesto que el número de posibles combinaciones es tan grande que (con la excepción de gemelos idénticos) son muy raros los emparejamientos exactos.

Los antígenos de histocompatibilidad difieren en su capacidad para provocar una respuesta inmunológica en un receptor injertado. En el ratón, un grupo de antígenos conocidos como sistema *H*-2 proporciona, con mucho, la barrera más fuerte para el trasplante. Si se intercambian injertos de piel entre ratones que tengan diferentes antígenos *H*-2 en sus células, el injerto resultará siempre rechazado en un plazo de 14 días. Injertos que difieran en antígenos de histocompatibilidad débiles, fuera del sistema *H*-2, pueden, sin embargo, requerir plazos de hasta 200 días para que se produzca el rechazo. Los antígenos *H*-2 fueron descritos por primera vez por Peter Gorer y sus colegas del Lister Institute of Preventive Medicine, en 1937, e investigados en profundidad por dicho autor y George Snell y sus colegas en el Jackson Laboratories en Bar Harbor, durante los 20 años siguientes.

Los fuertes antígenos *H*-2 del ratón tienen su contrapartida en los antígenos *HLA* del hombre, detectados en un principio sobre la superficie de los leucocitos sanguíneos a comienzos de la década de 1950. Aunque estos antígenos fueron más difíciles de definir por razones genéticas, se han deducido las características generales del sistema *HLA* mediante correlación de datos de un gran número de laboratorios aportados en una serie de reuniones internacionales de trabajo sobre tipaje de tejidos que se mantuvieron durante el final de la década de 1950 y los primeros años sesenta. Trabajos sobre estos antígenos han proporcionado una aplicación clínica directa en el emparejamiento de tejidos y órganos para la cirugía del transplante, así como para reducir la reacción inmunólogica del receptor contra el injerto. Además, la comprobación de que enfermedades como la esclerosis múltiple, la enfermedad de Hodgkin (cáncer de los ganglios linfáticos) y la hepatitis crónica se presentan con mayor frecuencia en individuos con determinados tipos de *HLA*, han provocado el planteamiento de cuestiones importantes sobre el origen genético de estas enfermedades y su relación con la función de los antígenos de histocompatibilidad.

Dado que los antígenos *H*-2 y *HLA* tienen muchas variantes genéticas, sería extremadamente difícil investigar su naturaleza molecular en poblaciones heterogéneas, tal como ocurre en el hombre. Los datos fundamentales sobre tales antígenos pueden explorarse más fácilmente en ratones, que, por intensos entrecruzamientos endogámicos, resultan genéticamente homogéneos. Por otra parte, muchos tests funcionales que no pueden hacerse con humanos sí se pueden realizar con ratones. Tal es la razón que explica que la mayor parte de este estudio se centre sobre los antígenos *H*-2.

Los antígenos *H*-2 y *HLA* son moléculas proteicas, que se adhieren firmemente a la superficie de casi todas las células. Con probabilidad, parte de la molécula se encuentra embebida en la membrana externa de la célula o la atraviesa.

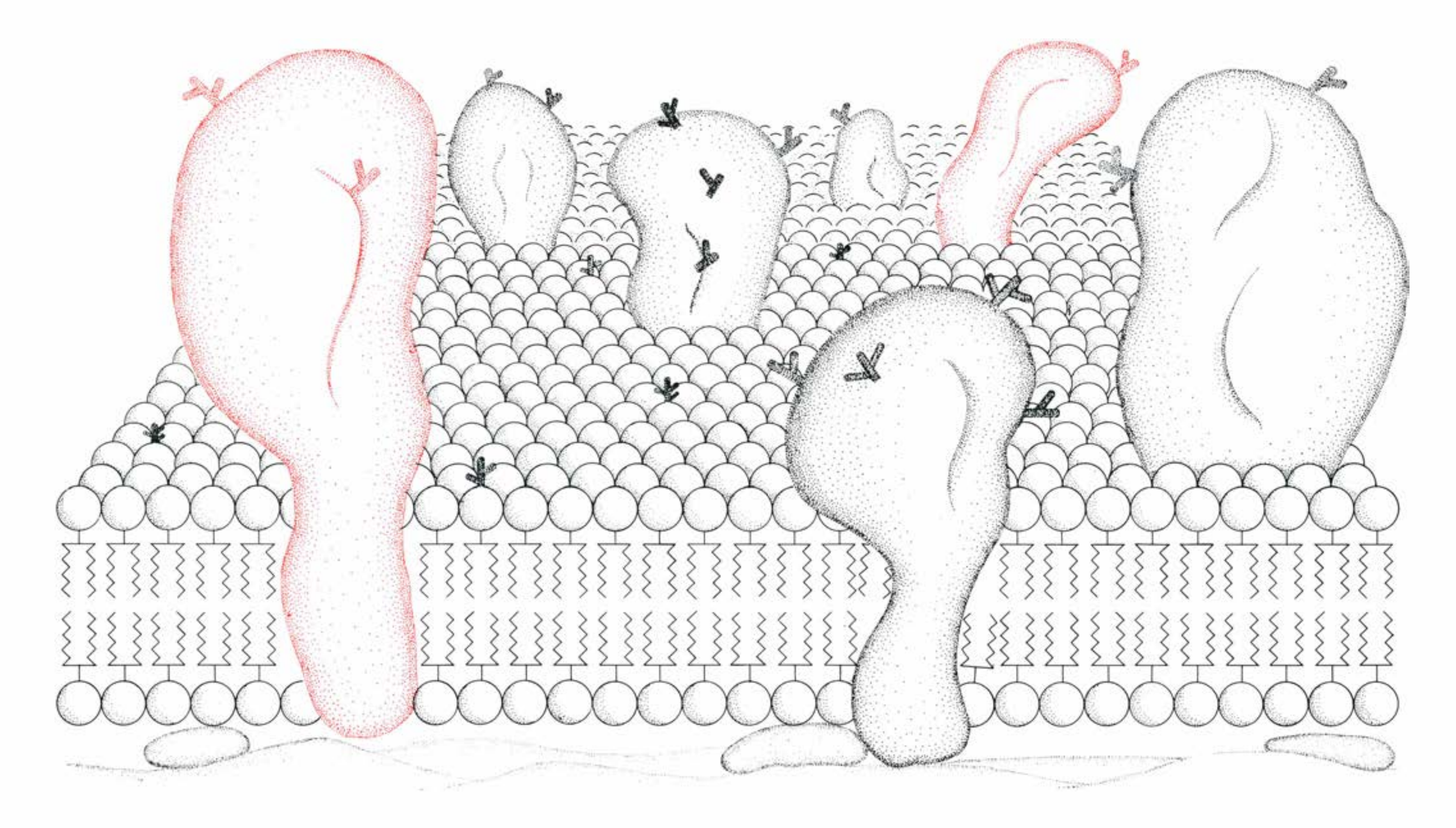
Como otras moléculas proteicas, los antígenos de histocompatibilidad constan de varias cadenas de aminoácidos; contienen, además, una pequeña cantidad de hidratos de carbono, por lo que se denominan glucoproteínas. Se detectan y clasifican por su capacidad para unirse a los anticuerpos específicos que se fabrican cuando se inyectan células de un animal en otro animal de la misma especie, genéticamente distinto. Algunas de las moléculas presentes en las células inyectadas resultan ser idénticas a las del receptor y otras, debido a las diferencias genéticas, son distintas. El receptor sintetizará anticuerpos específicos dirigidos contra las moléculas que son diferentes de las suyas.

Tales anticuerpos son los que definen la especificidad antigénica de las células inyectadas. Así, una determinada especificidad antigénica puede definir o reflejar la presencia, en las células del donante, de un antígeno ausente en las células del receptor, o bien puede reflejar diferencias estructurales entre variantes de un antígeno de histocompatibilidad que se encuentre presente tanto en las células del donante como en las del receptor.

Pueden obtenerse anticuerpos específicos para antígenos *H*-2 al inyectar a ratones de una raza isogénica con células de ratones de otra raza isógenica. Mediante cuidadosos entrecruzamientos pueden conseguirse razas que difieran solamente en una pequeña región de su mapa genético. Tales razas pueden utilizarse para producir anticuerpos específicos para un solo tipo de antígeno *H*-2. La aplicación sistemática de tales anticuerpos específicos, junto con la descripción del mapa genético y estudios genéticos de poblaciones, han capacitado a los inmunogenetistas para describir, detalladamente, los genes que codifican para los antígenos *H*-2. Estas investigaciones han revelado que los genes *H*-2 tienen muchas variantes y que se hallan estrechamente asociados con otros genes que desempeñan un papel fundamental en el funcionamiento del sistema inmune y en el desarrollo de embrión a adulto.

La región del material genético del ratón que codifica para los antígenos *H*-2 está constituida por un complejo de genes localizados a lo largo de un segmento del cromosoma 17, que es relativamente corto y que sólo representa una tresmilésima parte del ADN total del ratón. A pesar de todo, dicha región es lo suficientemente grande como para codificar hasta 200 proteínas de tamaño intermedio. Dos de esas proteínas son los antígenos *H*-2, que se codifican por genes localizados en ambos extremos del complejo *H*-2; se designan en el simbolismo *H*-2*K* y *H*-2*D*, respectivamente. Una célula somática del ratón contiene dos cromosomas 17, que recibe transmitidos uno por cada progenitor. En poblaciones heterogéneas, los genes de los progenitores serán diferentes, dando lugar a cuatro antígenos *H*-2 diferentes: dos *K* y dos *D*.

Se han descrito, al menos, 10 variantes de genes *H*-2*K* y 10 variantes de genes *H*-2*D* en cepas isogénicas de ratones. Las investigaciones realizadas por Jan Klein, de la University of Texas Southwestern Medical School, en Dallas, sobre poblaciones heterogéneas salvajes, de ratones, sugieren que el número de formas variables de cada gen puede superar el centenar. Las especificidades antigénicas para el *H*-2*K* y el *H*-2*D* son inmunológicamente distintas, pero ambos genes tienen muchos aspectos en común. Donald C. Shreffler y sus colaboradores, de la Facultad de Medicina de la Universidad de



PRINCIPALES ANTIGENOS DE HISTOCOMPATIBILIDAD, esquematizados aquí *(color)*; se hallan entre un vasto conjunto de proteínas embebidas en la bicapa lipídica que constituye la membrana externa de la célula. Una parte de la molécula de antígeno se extiende hacia el interior de la célula, en donde puede interactuar con proteínas citoplasmáticas, mientras que la otra parte sobresale justo por fuera de la célula y está disponible para interaccionar con otras células. Debido a la variabilidad genética de los antígenos de histocompatibilidad, éstos proporcionan a los tejidos de cada individuo su propia identidad química. Cuando se transplanta un trozo de piel o un órgano, de un individuo a otro, no emparentado, los linfocitos matadores del sistema inmunológico del receptor reconocen los antígenos de histocompatibilidad extraños.

INJERTO DE PIEL

RECEPTOR DE LA CEPA *A*

INJERTO ACEPTADO

RATON CEPA *A*

INJERTO DE PIEL

RECEPTOR DE CEPA *B*

10 A 14 DIAS

INJERTO RECHAZADO

INJERTO DE PIEL

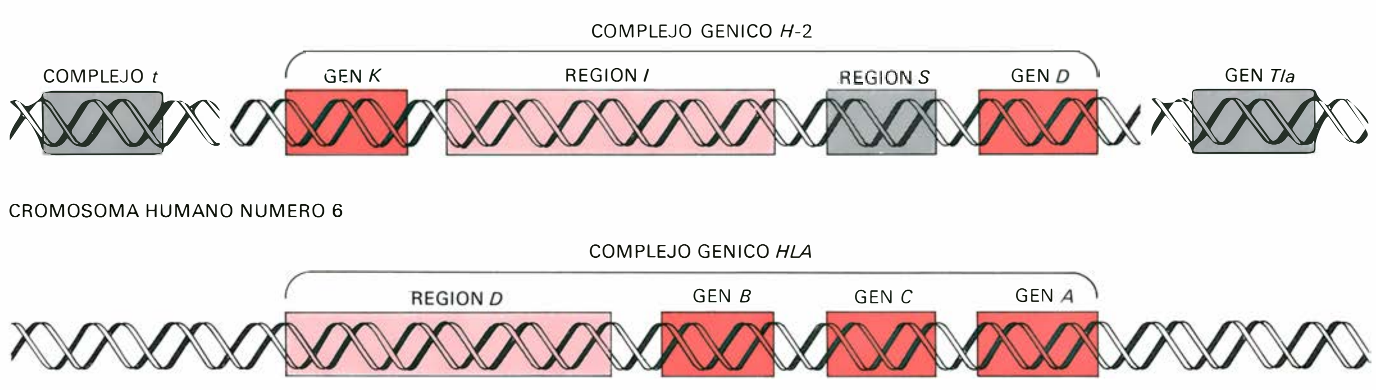
TRANSFERENCIA DE CELULAS LINFOIDES

RECEPTOR DE CEPA *B*

5 A 7 DIAS

RECHAZO ACELERADO

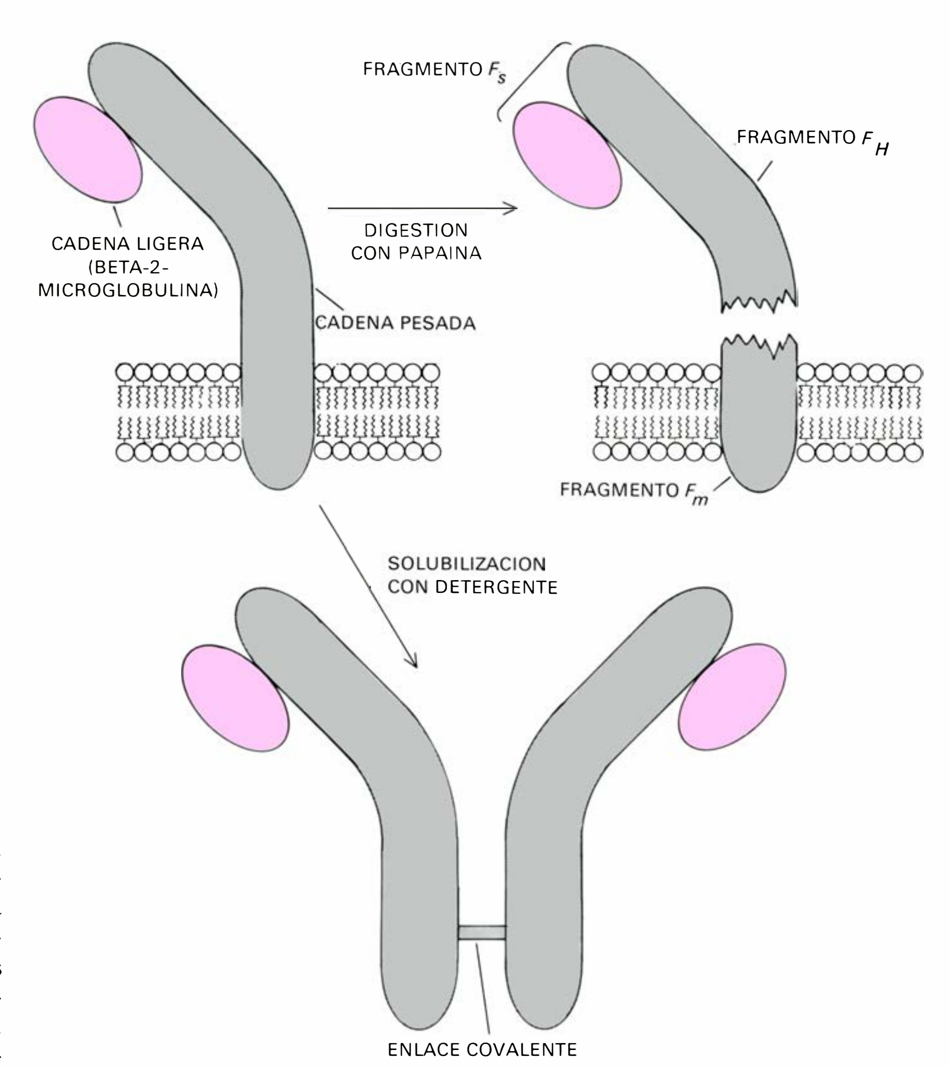
LA GENETICA DEL TRANSPLANTE se investiga mejor en los ratones, que pueden hacerse genéticamente homogéneos por medio de entrecruzamientos intensos de una misma cepa. Si la piel se transplanta desde un ratón de una cepa isogénica *(gris)* hasta otros de la misma cepa, el injerto prende indefinidamente. Si, por el contrario, la piel se transplanta a un miembro de una cepa genéticamente distinta *(color)*, el injerto se rechazará en el plazo de dos semanas. Cuando los linfocitos de un ratón que ya ha rechazado un injerto se inyectan en un ratón genéticamente idéntico al mismo tiempo que éste recibe un transplante del donante no emparentado, el injerto será rechazado con una rapidez doble. Este rechazo se debe a que los linfocitos transferidos estaban sensibilizados ya contra los antígenos de histocompatibilidad extraños.



REGIONES CROMOSOMICAS ESPECIFICAS codifican los antígenos de histocompatibilidad principales del ratón (los antígenos *H*-2) y del hombre (los antígenos *HLA*). El complejo *H*-2 *(arriba)* se encuentra localizado sobre el cromosoma número 17 del ratón. En ambos extremos del complejo se hallan los genes que especifican "fuertes" antígenos *H*-2*K* y *H*-2*D*. Entre estos genes se encuentran regiones genéticas relacionadas con la función del sistema inmune: la región *I*- que controla la fabricación de anticuerpos para ciertos antígenos extraños, y la región *S*, que controla la fabricación de un componente del sistema sérico del complemento, un sistema de proteínas que destruyen células antigénicas, una vez que éstas han reaccionado con los anticuerpos. Las regiones designadas *t* y *Tla* están localizadas por fuera del complejo *H*-2 pero pueden estar funcionalmente relacionadas con los genes *H*-2. El complejo *HLA (abajo)*, localizado a lo largo del cromosoma número 6 humano, contiene tres genes que codifican para antígenos de histocompatibilidad *(HLA-A, HLA-B, y HLA-C)* y una región *D* que es comparable a la región *I* del ratón. La topografía de los dos complejos génicos varía *(las regiones homólogas están coloreadas de igual manera).*

Michigan, han avanzado que ambos genes pueden haber evolucionado de copias duplicadas de un gen ancestral, las cuales, a través de mutaciones, dieron lugar progresivamente a diferentes especificidades proteicas.

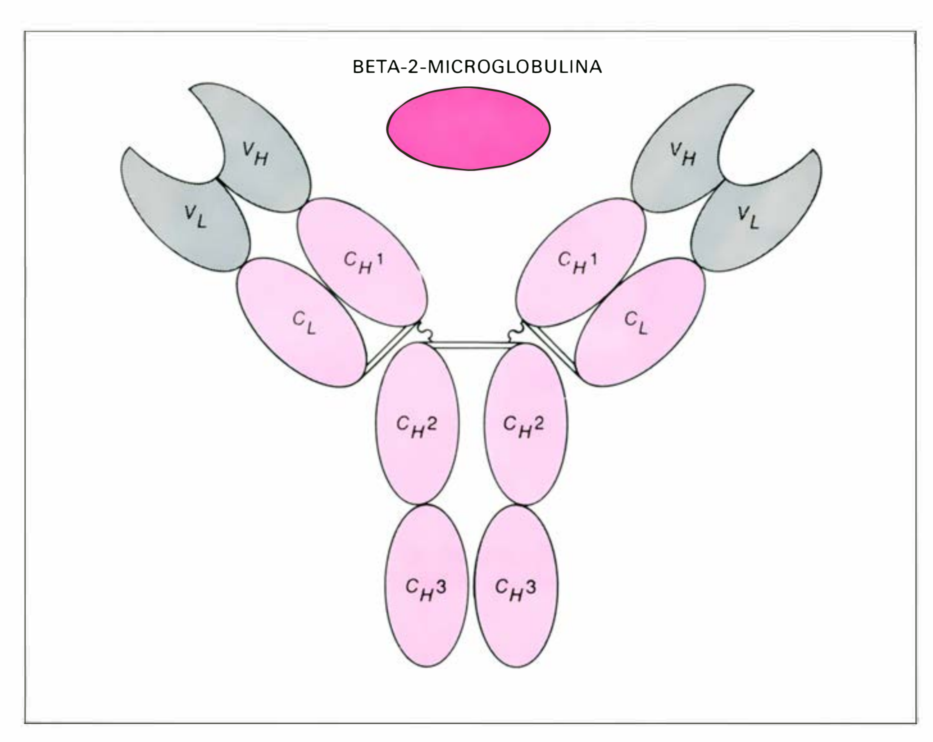
Otras partes del complejo *H*-2 que han podido ser identificadas abarcan la región *I* y la región *S*, las cuales se encuentran estrechamente relacionadas con el sistema inmunológico. La importancia de la región *I* se reconoció primero, cuando Baruj Benacerraf, que trabajaba en el National Institute of Allergy and Infectious Diseases, Hugh O. McDevitt, de la Stanford University Medical School, y Michael Sela, del Weizmann Institute of Science, mostraron que cierta razas de ratones isogénicos podían fabricar anticuerpos contra un determinado antígeno, mientras que otras razas no lo podían hacer. Subsiguientemente, hallaron que la capacidad para sintetizar anticuerpos estaba controlada por una serie de genes dentro de la región *I*, conocidos como genes *Ir* o genes de inmunorrespuesta. Aunque los productos de los genes *Ir* no se han identificado, sí se han podido detectar ciertas proteínas designadas por *Ia* (antígenos asociados a la región *I*) sobre la superficie de los linfocitos, que son los leucocitos sanguíneos involucrados en las reacciones inmunológicas. No obstante, todavía sigue sin saberse si los antígenos *Ia* vienen codificados por los genes *Ir*. La región *S*, que fue primeramente detectada por Shreffler y sus colegas, ha demostrado controlar la síntesis de un componente del complemento sérico; éste es el sistema de proteínas del suero sanguíneo que actúa para destruir células antigénicas, como bacterias y células injertadas, en cuanto hayan entrado en reacción con anticuerpos específicos.



LA ESTRUCTURA DE LOS ANTIGENOS *H*-2 difiere en función del modo en que se han extraído de la membrana celular. Cuando las moléculas *H*-2 se extraen con detergente, aparecen como un complejo de dos cadenas pesadas *(largas)* y dos cadenas polipeptídicas ligeras *(cortas)*. Las dos cadenas pesadas se encuentran unidas por un enlace covalente, pero las cadenas ligeras y pesadas se mantienen unidas entre sí sólo por débiles interacciones no covalentes. Cuando las moléculas *H*-2 se extraen de la membrana celular por acción de la enzima proteolítica papaina, se obtiene un simple fragmento designado F_s que consta de una parte de la cadena pesada (F_H) y de una cadena ligera completa. El fragmento más corto de la cadena pesada, designado F_m, permanece probablemente asociado con la membrana celular. En la superficie celular, no se encuentra la unión covalente entre las dos cadenas pesadas halladas en la solución de detergente. La estructura de los antígenos *H*-2 puede estar constituida por un complejo de una cadena pesada y una ligera, un complejo de dos cadenas ligeras y dos cadenas pesadas o bien una mezcla de ambas formas señaladas.

En resumen, el sistema *H*-2 es una región cromosómica única que transporta varios genes estrechamente relacionados todos ellos con el sistema inmune. Fuera del complejo, pero también a lo largo del cromosoma 17, existen al menos otras dos regiones que parecen estar relacionados con los genes *H*-2. Una de ellas es el gen *Tla*, que codifica para los antígenos timo-leucémicos *(TL)*, definidos por Edward A. Boyse, Lloyd J. Old y Elisabeth Stockert, del Sloan-Kettering Institute for Cancer Research (véase "Inmunología del cáncer", por Lloyd J. Old, INVESTIGACION Y CIENCIA, julio). Los antígenos *TL*, que se encuentran presentes sobre la superficie de células tímicas normales en algunas cepas de ratones, así como sobre las células leucémicas de casi todas las cepas, tienen muchos aspectos estructurales en común con los antígenos de histocompatibilidad *H-2K* y *H-2D*. Además, la expresión de los antígenos timoleucémicos sobre la superficie celular esta estrechamente relacionada con la expresión de los antígenos *H*-2. La segunda región relacionada con el complejo *H*-2 se define como complejo *t*; recuerda al complejo *H*-2 en que tiene muchas variantes genéticas. L. C. Dunn y Dorothea Bennett, de la Columbia University, demostraron, hace algún tiempo, que el complejo *t* representa un papel crítico en el desarrollo embrionario del ratón. Las mutaciones que ocurran en esta región pueden causar cambios visibles en la forma y tamaño de la cola del ratón; algunas mutaciones de estos genes resultan letales.

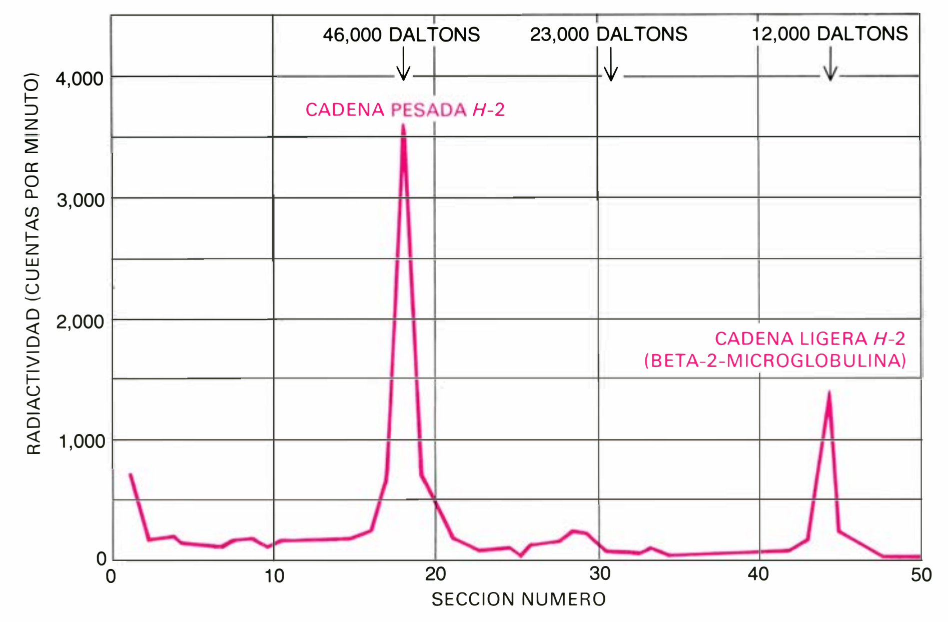
Para el complejo *HLA* humano se ha obtenido un mapa genético similar aunque menos detallado. La relativa carencia de conocimientos parte del hecho de que los anticuerpos humanos para la detección de antígenos *HLA* son bastante más difíciles de obtener de forma sistemática que los anticuerpos de ratón para el sistema *H*-2. Los anticuerpos anti-*HLA* se encuentran con frecuencia en el suero de mujeres que han tenido múltiples embarazos y han fabricado anticuerpos contra los antígenos *HLA* paternos



LA SEMEJANZA ESTRUCTURAL entre la beta-2-microglobulina (la cadena ligera componente de los antígenos *H*-2 y *HLA*) y las regiones constantes del anticuerpo *(color claro)* viene sugerida por las homologías en la secuencia de aminoácidos. Estos hallazgos implican que los antígenos de histocompatibilidad están relacionados genéticamente con los anticuerpos; plantean la cuestión de si la molécula *H*-2 está plegada en dominios funcionales semejantes a los de la molécula de anticuerpo.

expresados en las células del feto. Suelen obtenerse también anticuerpos contra dichos antígenos a partir de individuos que han recibido múltiples transfusiones sanguíneas y que, por tanto, han estado expuestos a antígenos *HLA* extraños presentes en los leucocitos de la sangre del donante, así como de voluntarios que han sido deliberadamente inmunizados con leucocitos de la sangre de otras personas.

El complejo *HLA* se encuentra localizado en el cromosoma 6 humano. Tiene muchos aspectos en común con el complejo *H*-2. Existen tres genes en el complejo que codifican para antígenos de histocompatibilidad, los cuales se han designado con los símbolos *HLA-A*, *HLA-B* y *HLA-C*. El *HLA-A* y el *HLA-B* parecen estar estechamente relacionados con el *H-2D* y el *H-2K* del ratón. No obstante, la topografía del complejo *HLA* difiere de la del complejo *H*-2 en que la región *HLA-D* (un área comparable a la región *I* del ratón) se localiza externamente a los genes *A* y *B*, mientras que la región *I* se encuentra entre el *H-2D* y el *H-2K*.



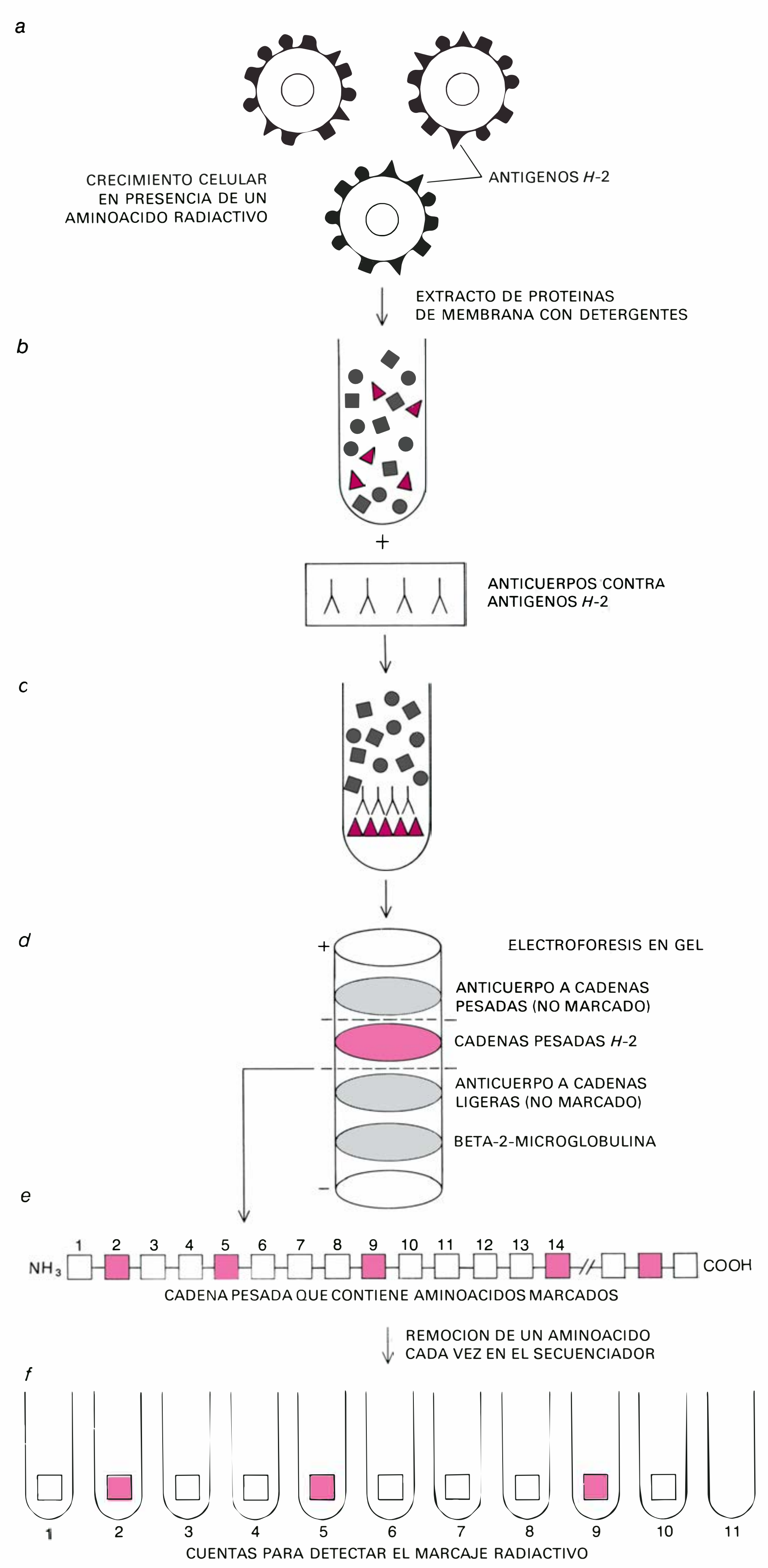
LA SEPARACION de las cadenas polipeptídicas pesadas y ligeras de los antígenos *H*-2 se logra por medio de electroforesis. Moléculas *H*-2 marcadas radiactivamente se aplican a la parte superior de un gel de poliacrilamida cilíndrico; el gel se somete a un campo eléctrico. Las cadenas pesadas y ligeras emigran hacia la base del gel, a diferentes velocidades según su peso molecular. El gel se corta en secciones finas y se mide la radiactividad de cada una de ellas. Así se obtienen dos picos de radiactividad correspondientes a las bandas purificadas de cadenas ligeras y pesadas del *H*-2.

Se conoce muy poco sobre la estructura molecular de los antígenos de histocompatibilidad. Ello se debe sobre todo a que, al igual que otras proteínas de la superficie celular, presentan problemas experimentales de importancia para el químico de proteínas: no se pueden obtener fácilmente en gran cantidad, no son solubles en agua ni en soluciones salinas sencillas y están entremezcladas con otras muchas moléculas sobre la superficie celular. A pesar de estas limitaciones, los antígenos *H*-2 resultan más accesibles para la realización de análisis estructurales que otras muchas glucoproteínas de la superficie celular; y esto por una serie de razones: primero, los numerosos estudios genéticos e inmunológicos de los antígenos *H*-2 proporcionan una base excelente para la investigación química; segundo, se pueden emplear anticuerpos para aislar antígenos *H*-2 específicos de mezclas complejas de proteínas; tercero, los ensayos para antígenos *H*-2 están bien establecidos y son fiables a pesar de no ser los ideales. Como resultados del análisis de la estructura molecular de los antígenos *H*-2, aparte de su valor para la comprensión del sistema de histocompatibilidad, proporcionan datos inapreciables para el análisis futuro de otras glucoproteínas de la superficie celular. Junto con Gerald M. Edelman, Roland Henning, Robert J. Milner y Konrad Reske, de la Rockefeller University, he venido empleando una variedad de métodos para la caracterización de proteínas con el fin de obtener un modelo más preciso de la estructura de los antígenos *H*-2, tanto en solución como en la superficie celular.

La molécula de antígeno *H*-2 está constituida por un complejo de cadenas polipeptídicas, consistentes cada una de ellas en una secuencia lineal de aminoácidos unidos por enlaces peptídicos. Se han identificado dos tipos de cadenas polipeptídicas constituyentes: cadenas pesadas (largas) y cadenas ligeras (cortas). Como ha demostrado Stanley Nathenson y sus

colegas, de la Facultad de Medicina Albert Einstein de Nueva York, las moléculas *H*-2 pueden obtenerse para estudios estructurales, bien por disrupción de la membrana celular con un detergente o bien por el empleo de la enzima proteolítica papaína para extraer químicamente de la superficie celular la molécula.

Al extraer la molécula *H*-2 con un detergente, se mantendrá en solución solamente si el detergente se encuentra presente. Nosotros hemos hallado bajo estas condiciones que la molécula *H*-2 está constituida por dos cadenas pesadas idénticas y dos cadenas ligeras idénticas. Las dos cadenas pesadas se mantienen unidas merced a un enlace covalente, por lo menos. Las cadenas ligeras se mantienen unidas a las pesadas solamente por enlaces más débiles, no-covalentes, por ejemplo, puentes de hidrógeno y fuerzas electrostáticas. Cuando la molécula se extrae con papaína, se libera un fragmento hidrosoluble de ella, denominado F_s. Un fragmento similar se libera cuando el material detergente solubilizado se trata con papaína. En ambos casos, el fragmento F_s consta de una cadena ligera completa asociada a un fragmento (F_H) de la cadena pesada. Una porción más pequeña de la cadena pesada y designada F_m permanece probablemente asociada a la membrana celular después de que la cadena pesada se haya roto por acción de la papaína. No se ha podido aislar todavía este fragmento más pequeño, pero posiblemente comprenda la región de la cadena pesada que interacciona con la membrana junto con las subunidades aminoacídicas que forman los enlaces covalentes entre dos cadenas



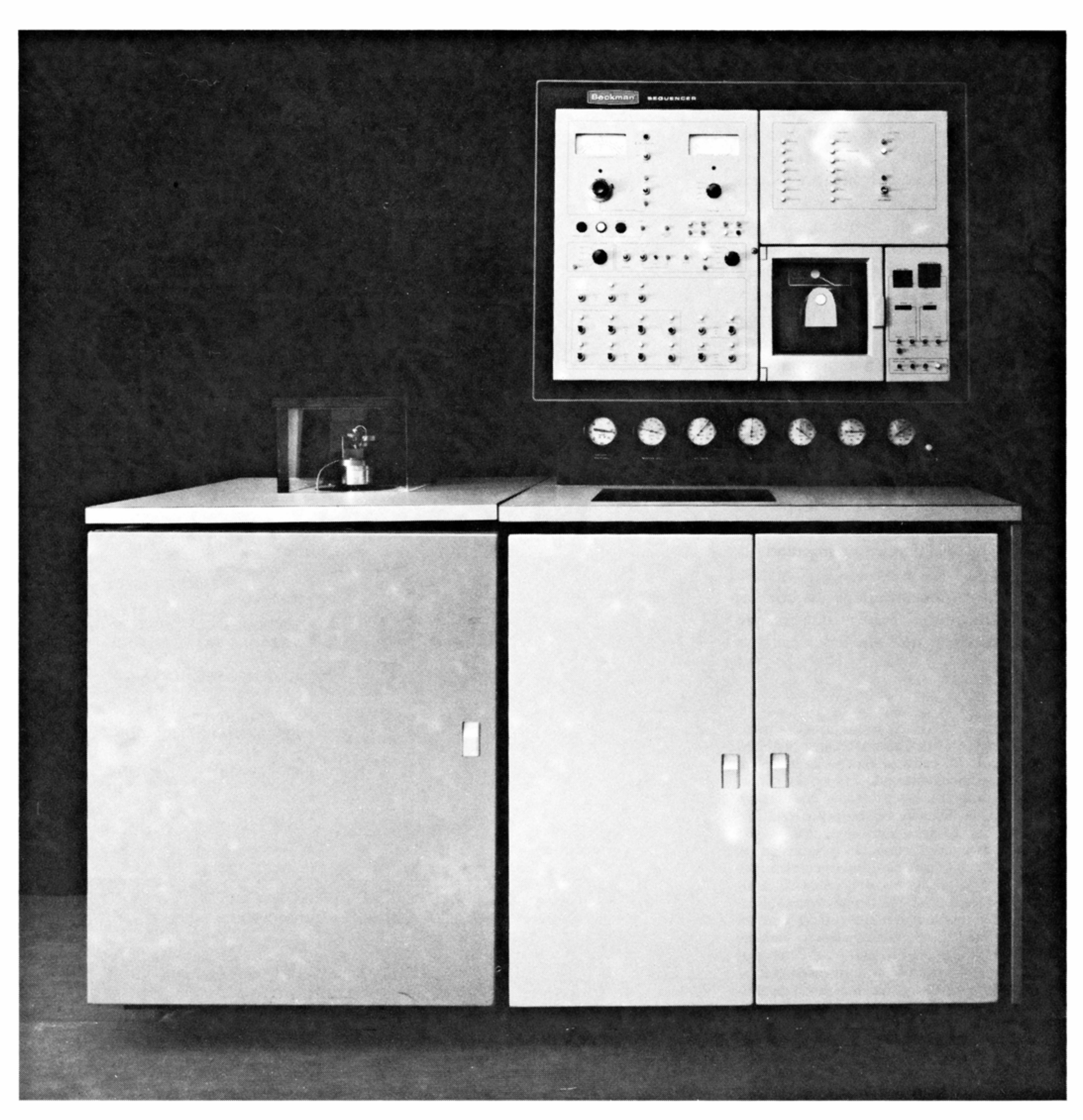
LA SECUENCIACION de las cadenas pesadas *H*-2 requiere una nueva técnica analítica microquímica. Se cultivan células de ratón en presencia de un aminoácido radiactivo que se incorpora en todas las proteínas nuevamente sintetizadas, incluyendo los antígenos *H*-2 *(a)*. Las proteínas de membrana marcadas se van extrayendo con detergentes *(b)*, y se aíslan los antígenos *H*-2 por precipitación con anticuerpos específicos *(c)*; luego, las moléculas de antígeno se redisuelven en sus componentes (cadenas ligeras y pesadas) por electroforesis en gel de poliacrilamida *(d)*. Las cadenas pesadas purificadas se eluyen del gel y se determina la posición en que aparece el aminoácido radiactivo a lo largo de la cadena eliminando un aminoácido cada vez de la proteína, con la ayuda de una máquina automática llamada "secuenciador" *(e)*. Finalmente, se mide la radiactividad de los aminoácidos eliminados secuencialmente, para detectar la presencia o ausencia de la etiqueta radiactiva *(f)*. Todo el proceso debe repetirse muchas veces para determinar la localización a lo largo de la cadena polipeptídica de cada uno de los veinte aminoácidos de que están formadas las moléculas proteicas.

pesadas cuando se encuentran en una solución de detergente.

Las moléculas *H*-2 presentes en las células difieren de las que se encuentran en la solución detergente en que aquéllas carecen de los puentes covalentes entre las cadenas pesadas. Estos hallazgos sugieren las siguientes posibilidades para la estructura de las moléculas *H*-2 en la superficie celular: primero, un complejo consistente en una cadena ligera y otra pesada: segundo, un complejo constituido por dos cadenas ligeras y dos cadenas pesadas; y tercero, una combinación de ambos. Estas posibilidades son relevantes desde un punto de vista funcional, ya que implican que los antígenos presentes en las células sean susceptibles de interacción con otras moléculas, del mismo modo que interactúan entre sí en una solución detergente.

Recientes investigaciones han identificado la cadena polipeptídica ligera del *H*-2, *HLA* y *TL* como la beta-2 microglobulina, una proteína que ya les era conocida a los bioquímicos. Esta molécula no incorpora hidratos de carbono y no parece poseer ninguno de los determinantes antigénicos que son característicos de un definido tipo *H*-2 o *HLA*. Por tanto, los hidratos de carbono y las especificidades antigénicas deben correr a cargo de las cadenas pesadas.

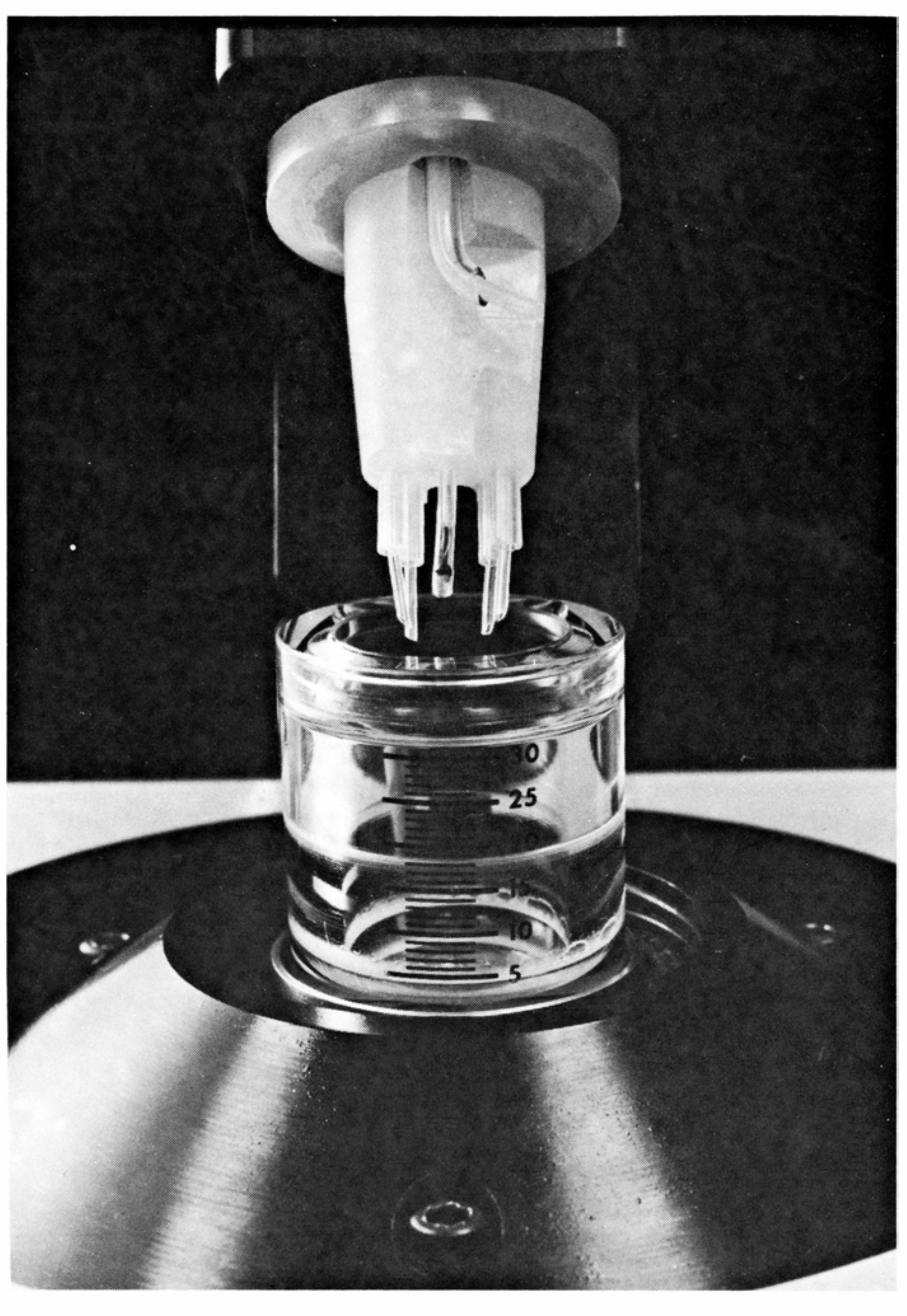
La beta-2-microglobulina humana fue descubierta por Ingemar Berggard y Alexander G. Bearn, de la Rockefeller Uni-



SECUENCIADOR AUTOMATICO, del laboratorio del autor en la Universidad de Rockefeller. Puede determinar la secuencia de aminoácidos de una cadena polipeptídica de hasta 50 aminoácidos consecutivos. La máquina usa unas condiciones de control muy cuidadas para llevar a cabo una reacción que remueve secuencialmente un aminoácido cada vez de la cadena. La secuencia está sujeta a programación controlada.

versity, en 1968, en la orina de pacientes con disfunción renal. Desde entonces, se ha estudiado detalladamente esta proteína y ha quedado de manifiesto su síntesis por casi todas las células del organismo. También se encuentra presente en otras especies; así, en el ratón, el perro, el conejo, el cobayo y la rata. En contraste con la variabilidad de las cadenas pesadas de los antígenos de histocompatibilidad, no se han detectado variantes de la beta-2-microglobulina dentro de una especie, ni a través del empleo de métodos químicos ni mediante el uso de métodos inmunológicos. Las secuencias aminoácidas de la beta-2-microglobulina obtenidas de diversas especies se han determinado parcialmente; tan sólo muestran pequeñas diferencias en relación con el número que cabría esperar para proteínas funcionalmente semejantes de diferentes especies. Esta semejanza sugiere que la estructura de la beta-2-microglobulina en el hombre es representativa de la estructura de la molécula en todas las demás especies de vertebrados, incluido el ratón.

El dato más sorprendente de la beta-2-microglobulina es su semejanza a ciertas porciones de las moléculas de los anticuerpos. Desde hace algún tiempo se sabe que las cadenas polipeptídicas ligeras y pesadas que constituyen las moléculas de anticuerpo tienen dos tipos de regiones: regiones variables que difieren en la secuencia de aminoácidos de una molécula a otra, y regiones constantes que son invariables en su secuencia de aminoácidos para una clase de cadenas pesadas o cadenas ligeras. Tres segmentos en la región constante de la cadena pesada del anticuerpo se parecen entre sí y a la región constante de la cadena ligera en tamaño, estructura general y secuencia aminoácida. Sobre las bases de estas observaciones, Edelman propuso que tanto las cadenas polipeptídicas ligeras como pesadas del anticuerpo estaban plegadas en dominios compactos correspondientes a las regiones constantes y variables. Esta hipótesis se ha confirmado actualmente mediante análisis estructurales utilizando la técnica de cristalografía de rayos X (véase "El lado de unión de los anticuerpos", por J. Donald Capra y Allen B. Edmundson; INVESTIGACION Y CIENCIA, marzo). En suma, estos datos, sugieren con las pruebas genéticas, que los genes que codifican para moléculas de anticuerpos se desarrollan a través de la duplicación de un gen ancestral lo suficientemente largo como para codificar un dominio simple de anticuerpo, un polipéptido de unos 100 aminoácidos de longitud.



EN PRIMER PLANO, la cámara de reacción del secuenciador. La taza de vidrio gira durante la secuencia de la reacción y se mantiene a una temperatura constante por medio de un elemento de calentamiento. Los tubos de plástico del interior del cabezal de distribución *(que aquí está elevado para su visualización)* **permite la introducción de soluciones y nitrógeno en el interior de la cámara. Al comienzo de la reacción, se introduce la muestra polipeptídica como una película sobre la superficie interna del vaso giratorio. El reactivo fenilisotiocianato puede entonces copularse con el extremo de las cadenas polipeptídicas, y se elimina un aminoácido de las cadenas por exposición del material a condiciones acídicas. Finalmente, el aminoácido liberado se extrae de la mezcla de reacción con un disolvente orgánico, y se aspira del vaso giratorio a través del tubo en forma de** *L* **sobre el lado del cabezal de distribución** *(arriba).* **La secuencia de reacción comienza de nuevo para remover el aminoácido siguiente de la cadena. Los aminoácidos que se han ido eliminando se secuencian posteriormente con la utilización de una amplia variedad de técnicas cromatográficas.**

La cadena polipeptídica de la beta-2-microglobulina está constituida también por unos 100 aminoácidos, y recuerda al dominio constante de los anticuerpos en muchos de sus aspectos estructurales. Investigaciones llevadas a cabo por Oliver Smithies, de la Universidad de Wisconsin, en colaboración con Dave Poulik, de la Wayne State University, así como nuestra determinación de la secuencia completa de aminoácidos de la beta-2-microglobulina humana, han demostrado que se encuentra tan estrechamente relacionada con los dominios constantes de an-

ticuerpos como dichos dominios entre sí. Sobre las bases de estos datos, hemos concluido que el gen para la beta-2-microglobulina se desarrolló del mismo gen ancestral que dio lugar a los genes para las moléculas de anticuerpo.

El hecho de que la beta-2-microglobulina se halle asociada a los antígenos *HLA* y *H*-2 sugiere una posible relación evolutiva entre las moléculas de anticuerpo y las moléculas de antígeno de histocompatibilidad. Además, se plantea la cuestión de si las cadenas polipeptídicas ligeras y pesadas de los antígenos *HLA* y *H*-2 se encuentran plegadas en dominios funcionales semejantes a los que se hallan en las moléculas de anticuerpos. Las contestaciones a estas preguntas vendrán dadas por la determinación de la secuencia de aminoácidos de las cadenas pesadas *H*-2 y por el análisis de cristalografía de rayos X de la estructura tridimensional de la beta-2-microglobulina; ambos aspectos están siendo investigados actualmente en nuestro laboratorio.

El análisis de la secuencia de aminoácidos de las cadenas pesadas resulta bastante más difícil que el de la beta-2-microglobulina. Por una parte, la cadena pesada *H*-2 debe extractarse intacta de la célula con el detergente, y mantenerse en dicha solución. Por otra, con el fin de obtener una cantidad de cadenas pesadas *H*-2, suficiente para determinar su secuencia de aminoácidos mediante técnicas convencionales de secuenciación, es necesario un gran número de células; por lo que dicho intento se hace demasiado costoso para que resulte práctico. Nosotros y otros autores que trabajábamos en este campo, desarrollamos un método microquímico para determinar la secuencia de aminoácidos de una proteína cuando se cuenta con mínimas cantidades de ella.

El método comprende la incorporación de moléculas de un aminoácido dado, marcadas radiactivamente, en las cadenas pesadas *H*-2 con el fin de identificar las posiciones de dicho aminoácido a lo largo de la cadena polipeptídica. Hemos cultivado células espléncias de ratón o células de tumores de ratón en un medio que contiene un aminoácido –por ejemplo la tirosina– que se había marcado por la substitución de los átomos de hidrógeno en su molécula por átomos de tritio, que es el isótopo radiactivo del hidrógeno. La tirosina marcada con tritio queda captada por las células e incorporada dentro de todas las proteínas sintetizadas de nuevo, incluyendo los antígenos *H*-2 que constantemente se van liberando y resintetizando. Después, se rompen las células y se solubilizan las proteínas de la membrana con detergente. Posteriormente, se añaden anticuerpos específicos para los antígenos *H*-2 a la mezcla de proteínas solubilizadas con el fin de recuperar las moléculas de *H*-2 en forma de precipitado.

Este precipitado de antígenos *H*-2 unidos a anticuerpos específicos se concentra mediante centrifugación de la solución. Luego se resolubiliza en detergente, separándolo en sus componentes moleculares mediante electroforesis de alta tensión en gel de poliacrilamida. En este proceso, las cadenas ligeras y pesadas *H*-2 se separan y migran a diferentes velocidades a través del gel, separándose en finas bandas según su tamaño. La región de gel que contiene la banda de cadena pesada *H*-2 marcada radiactivamente, se aísla, y las moléculas se eluyen del gel. Los productos finales de la purificación son cadenas pesadas *H*-2 que encierran una subunidad de tirosina

POSICION DE LOS AMINOACIDOS A LO LARGO DE LA CADENA PESADA

	ANTIGENOS PRINCIPAL.	VARIANTES GENETICAS					5					10					15					20					25		
RATON	*H-2K*	K^b	■	Pro	His		Leu	Arg	Tyr	Phe	Val	Thr	Ala	Val		Arg	Pro		Leu			Pro	Arg	Tyr				Leu	Tyr
		K^d	Met	■	His		■	Arg	Tyr	■	■	Thr		■		Arg	Pro		■			Pro	Arg	Phe					Tyr
		K^k	Met	Pro	His		Leu	Arg	Tyr	Phe	His		Ala	Val		Ile	Pro		Leu		Lys	Pro	Phe	Ala					Tyr
	H-2D	D^b		Pro				Arg	Tyr				Ala	Val		Arg	Pro		Leu			Pro	Arg	Tyr					Tyr
		D^d	Met	■	His		Leu	Arg	Tyr	Phe	Val	Thr	Ala	Val	Thr	Arg	Pro		■			Pro	Arg	Tyr					Tyr
HOMRRE	*HLA-A*	*A2*	Gly	Ser		Ser	Met	Arg	Tyr	Phe	Phe	Thr	Ser	Val	Ser	Arg	Pro	Gly		Gly	Glu			Phe	Ile	Ala	Val		
		A1,2		Ser		Ser	Met	Arg	Tyr	Phe	Phe	Thr	Ser	Val	Ala	Arg	Pro	Gly											
	HLA-B	*B7*	Gly	Ser		Ser	Met	Arg	Tyr	Phe	Tyr	Thr	Ser	Val	Ser	Arg	Pro	Gly		Gly	Glu			Phe	Ile	Ala	Val		
		B8,13		Ser		Ser	Met	Arg	Tyr	Tyr	Tyr	Ser	Ala	Val	Ser	Arg	Pro	Gly											

Ala ALANINA
Arg ARGININA
Glu ACIDO GLUTAMICO
Gly GLICINA
His HISTIDINA
Ile ISOLEUCINA
Leu LEUCINA
Lys LISINA
Met METIONINA
Phe FENILALANINA
Pro PROLINA
Ser SERINA
Thr TREONINA
Tyr TIROSINA
Val VALINA

□ LOS AMINOACIDOS DE ESTAS POSICIONES NO HAN SIDO DETERMINADOS
■ (rosa) AMINOACIDOS COMPARTIDOS POR LOS *H*-2 Y LOS *HLA*
■ (gris) AMINOACIDOS COMPARTIDOS POR LOS *H-2K* Y LOS *H-2D*
■ (rosa claro) AMINOACIDOS COMPARTIDOS POR LOS *HLA-A* Y LOS *HLA-B*
■ AMINOACIDOS NO DETERMINADOS PERO QUE SE HAN VISTO EN ESTA POSICION EN OTRAS CADENAS

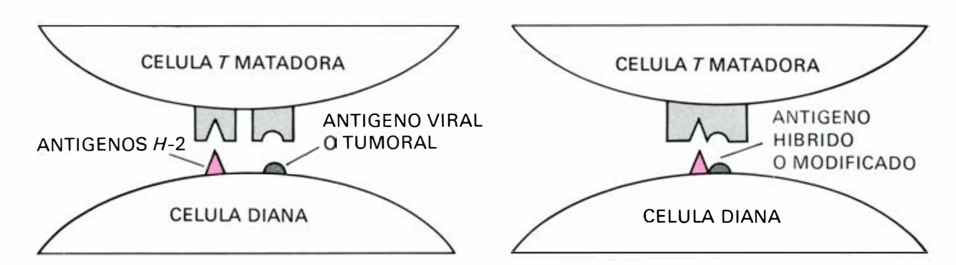
SECUENCIAS PARCIALES DE AMINOACIDOS de las cadenas pesadas de los antígenos *H*-2 y *HLA*. Revelan amplias regiones de semejanza; las diferencias pueden reflejar la variabilidad genética generalizada de estas moléculas. Las diferencias registradas en la secuencia se encuentran bien distribuidas a lo largo de los segmentos de la cadena, lo que sugiere que las distintas especificidades antigénicas se hallan determinadas por largas regiones de la molécula. En los anticuerpos, la secuencia de aminoácidos variables se limita a un segmento de cada cadena.

radiactiva en cada una de las posiciones en donde la tirosina aparece normalmente en la secuencia de aminoácidos. Las posiciones de otros aminoácidos a lo largo de la cadena pueden detectarse también mediante la incubación de las células en un medio que contenga moléculas de cada aminoácido marcadas con tritio.

Las posiciones exactas en que aparece el aminoácido marcado se determina con la ayuda de una máquina automática, denominada secuenciador, que elimina un aminoácido cada vez por medios químicos, comenzando por el extremo terminal amino (NH_2) de la cadena polipeptídica.

La reacción química requerida fue desarrollada por Pehr Edman, de la Universidad de Lund, quien, en colaboración con Geoffrey Begg, del St. Vincent's Hospital for Medical Research, en Melbourne, Australia, ideó subsiguientemente un sistema automático para controlar con precisión las condiciones de la reacción. El rendimiento de la liberación de cada aminoácido de la cadena supera el 95 por ciento; las secuencias aminoácidas pueden determinarse para un total de 30 a 50 subunidades de aminoácidos consecutivos, lo cual es mucho más de lo que se puede estudiar de una sola vez por medio de los métodos analíticos manuales. Habitualmente, los derivados de los aminoácidos liberados por la reacción de Edman se identifican por una serie de técnicas de separación cromatográfica. Debido a las pequeñas cantidades de proteína con que se trabaja en estos estudios, es imprescindible el uso de aminoácidos marcados radiactivamente.

Por medio de este método microquímico podemos detectar los aminoácidos de la molécula de la cadena pesada *H*-2 a un nivel de unas 10^8 moléculas, en tanto que el secuenciamiento convencional requiere casi un millón más de moléculas para que la detección sea adecuada (10^{14}). El problema más importante de esta metodología es que requiere mucho tiempo para una exacta identificación de cada aminoácido marcado, puesto que sólo se puede trabajar con uno o dos aminoácidos en cada prueba. Debe disponerse, pues, una preparación separada de cadenas *H*-2 pesadas purificadas, para determinar la localización, a lo largo de la cadena polipeptídica, de cada uno de los 20 aminoácidos que la constituyen. Además, algunos aminoácidos no pueden detectarse porque se incorporan en la proteína en cantidades insuficientes, ya sea porque no se transportan eficazmente al interior de la célula o porque las moléculas marcadas radiactivamente están muy diluidas en un gran acervo de moléculas similares, pero no marcadas, que se hallan en el interior de la célula. Nuestras secuencias adolecen, pues, de una serie de deficiencias.



MODELOS ALTERNATIVOS, que explican la observación de que los linfocitos denominados células *T* matadoras, las cuales pueden destruir selectivamente células del cuerpo infectadas por virus o cancerosas, tienen que detectar al menos una molécula de antígeno *H*-2 en la superficie de la célula diana antes de que ocurra la muerte celular. El proceso de reconocimiento puede comprender dos moléculas receptoras separadas o receptoras sobre la superficie de la célula *T* matadora (una para el antígeno *H*-2 y otra para el vírico). Alternativamente, un receptor sobre la célula matadora podría reconocer una estructura híbrida formada por la asociación de los dos antígenos.

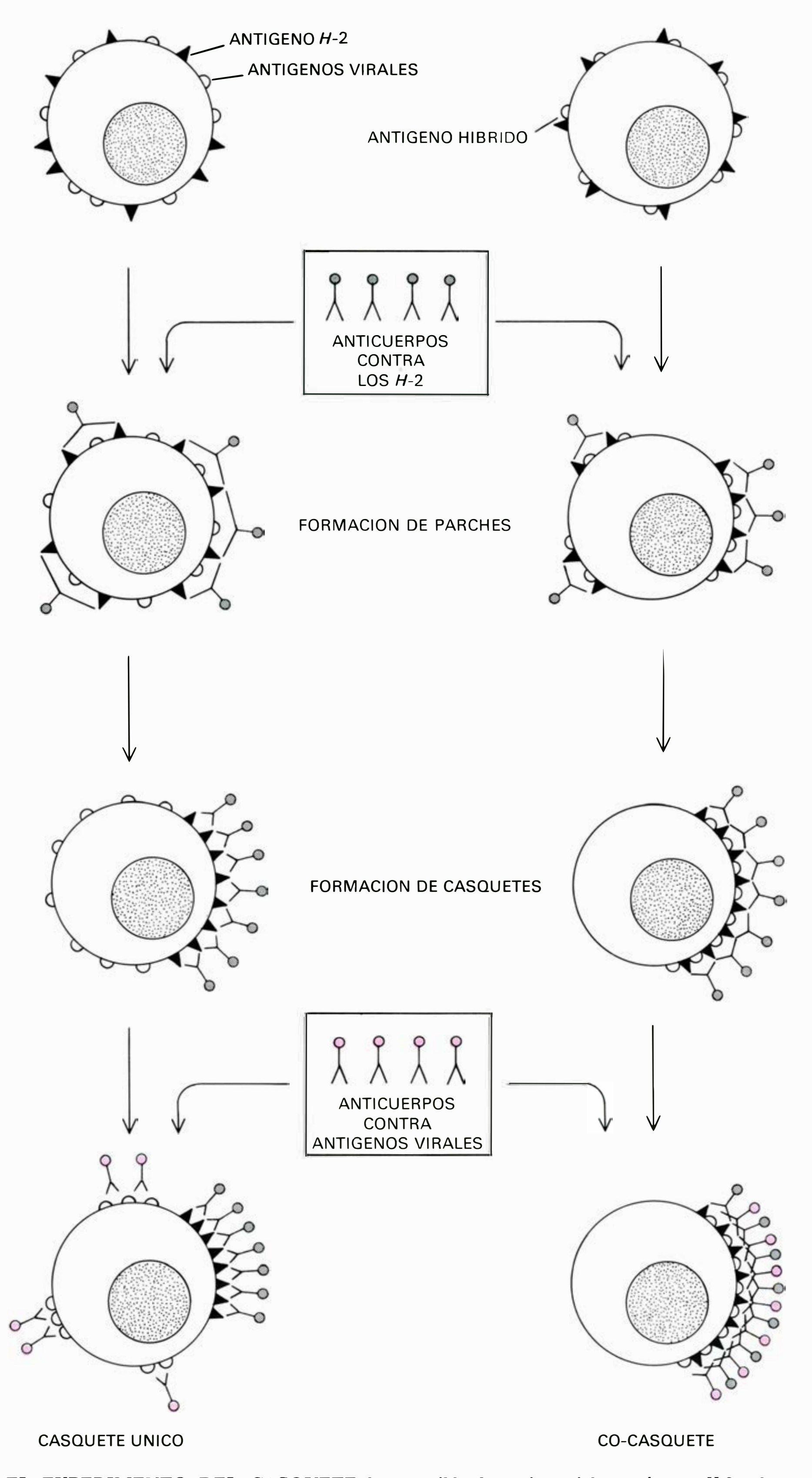
A pesar de estas limitaciones, durante los dos últimos años hemos obtenido la secuencia de los 27 primeros aminoácidos de varias cadenas pesadas *H*-2, al igual que lo han hecho Jack Silver y Leroy E. Hood, del Instituto de Tecnología de California, Nathenson y sus colaboradores, y un grupo encabezado por Jonathan W. Uhr, de la Facultad de Medicina de la Universidad de Texas Southwestern. Aunque los datos son todavía limitados, resultan ya suficientes para proporcionarnos una valiosa información sobre el tipo de variación genética que se halla en las cadenas pesadas, así como sobre el origen de los genes *H*-2*K* y *H*-2*D*.

Las secuencias aminoacídicas de todos los antígenos *H*-2 examinados hasta ahora son muy similares. Los productos de los genes *H*-2*K* se parecen de tal manera a los productos de los genes *H*-2*D* que no hay característica alguna en la secuencia de aminoácidos que permita distinguir un tipo de antígeno del otro. Estos resultados apoyan la hipótesis de Shreffler y de Klein en torno al gen ancestral común, a partir del cual se derivarían los genes *H*-2*K* y *H*-2*D*, y también la hipótesis de que todos los genes que se hallan en el complejo *H*-2 están estrechamente relacionados entre sí.

Las secuencias de aminoácidos en las variantes genéticas de las cadenas pesadas *H*-2 son similares, pero pueden reconocerse algunas diferencias características. Por ejemplo, los antígenos *H*-2 pueden tener aminoácidos distintos en las posiciones 1, 2, 8, 9, 14, 17, 22 y 23. Todavía no hemos podido correlacionar las diferencias en las secuencias con las distintas especificidades antigénicas de las variantes *H*-2, pero es evidente que ambas propiedades se relacionan entre sí. Las diferencias en la secuencia de aminoácidos parecen estar bien distribuidas a lo largo de dicha secuencia y no se confinan a un segmento específico corto. Por tanto, las distintas especificidades antigénicas que reflejan la variabilidad genética de los antígenos *H*-2 parecen generarse en las substituciones de aminoácidos presentes en una porción considerable de las cadenas pesadas *H*-2.

Una importante cuestión que planteó el estudio de la beta-2-microglobulina es si la cadena pesada *H*-2 es o no similar, en su secuencia de aminoácidos, a las regiones constante o variable de las moléculas de anticuerpo. Hemos comparado nuestras secuencias de cadenas pesadas *H*-2 con las de varias moléculas de anticuerpo, pero no hemos hallado semejanzas convincentes. Los datos sobre las cadenas pesadas *H*-2 son aún limitados y están circunscritos a una pequeña región cercana al extremo terminal amino de la molécula, de suerte que tendremos que esperar a disponer de secuencias más completas antes de poder emitir un juicio definitivo al respecto.

Paralelamente al trabajo llevado a cabo sobre las cadenas pesadas *H*-2 se está haciendo un intenso esfuerzo para determinar la estructura de las cadenas pesadas *HLA*. Jack L. Strominger y sus colegas de la Universidad de Harvard y un grupo de investigadores británicos, de los laboratorios de John Bridgen de la Universidad de Cambridge, Walter Bodmer, de la Universidad de Oxford, y Michael Crumpton, del National Institute for Medical Research at Mill Hill, han estudiado el problema desarrollando cultivos de grandes cantidades de células humanas, procedentes de líneas dotadas de una concentración anormalmente ele-



EL EXPERIMENTO DEL CASQUETE hace posible determinar si los antígenos *H*-2 y los antígenos víricos se encuentran estrechamente asociados sobre la superficie de las células tumorales. Primero, se marcan los anticuerpos específicos para los antígenos *H*-2 y los específicos para los antígenos víricos con colorantes, fluoresceína y rodamina, que rinden coloraciones verde y roja, respectivamente, bajo el microscopio de fluorescencia. Después, se añaden a las células anticuerpos contra los antígenos *H*-2 que, al unirse a éstos en la superficie celular, forman parches y, finalmente, forman un casquete en un polo de la célula. Luego se fijan las células y se añaden los anticuerpos contra los antígenos víricos. Si los dos antígenos son independientes, los antígenos *H*-2 constituirán un casquete pero los antígenos víricos permanecerán esparcidos en la superficie celular *(a)*. Si los dos antígenos en cuestión están asociados formarán un casquete juntos *(b)*.

vada de antígenos *HLA* en su superficie. Todos estos grupos de investigación han conseguido obtener una cantidad suficiente de cadenas pesadas *HLA*, que permite determinar la secuencia aminoacídica parcial de éstas sin necesidad de incorporar aminoácidos radiactivos en las cadenas polipeptídicas.

La comparación de las secuencias *H*-2 y *HLA* obtenidas hasta el momento satisface la expectación de que serían similares. Este hallazgo es muy significativo por dos razones. En primer lugar, apoya la hipótesis de que los principales antígenos de histocompatibilidad del retón y del hombre se asemejan entre sí y tienen, probablemente, la misma función biológica. Las investigaciones que se realicen sobre el sistema *H*-2 del ratón serán, pues, aplicables al estudio del sistema *HLA*. En segundo lugar, las semejanzas en las secuencias del *H*-2 y del *HLA* dan validez a la técnica microquímica que hemos empleado para el análisis secuencial de la estructura de los antígenos *H*-2, ya que las secuencias *HLA* fueron determinadas por medio de una metodología completamente diferente. Esta verificación es importante por cuanto muestra que el método resulta apropiado para el estudio de cualquier proteína de la superficie celular contra la cual puedan prepararse anticuerpos. Todavía nos queda el reto de la extensión de este sistema a la secuenciación de los aminoácidos de regiones mayores de la molécula de *H*-2. Por medio de la determinación de la secuencia de una porción importante de la cadena pesada del *H*-2, esperamos distinguir las características de la molécula de *H*-2 que inciden en la comprensión de su función biológica.

Como ya he indicado al comienzo de este artículo, los antígenos de histocompatibilidad no pueden haber evolucionado con el único fin de aumentar las dificultades del cirujano de transplante. ¿Qué podemos decir, pues, acerca de su función biológica en el huésped? Las explicaciones más prometedoras vendrán de la consideración de los antígenos *H*-2 en su papel de defensa del organismo contra las anormalidades que potencialmente pudieran aparecer sobre sus propias células. De acuerdo con la teoría de la vigilancia inmunológica, en cada momento se están produciendo en el organismo mutaciones deletéreas, capaces de causar el cáncer y otras disfunciones celulares. Dichas mutaciones determinan cambios en la estructura de la superficie de las células que luego reconoce el sistema inmunológico, especialmente a través de una clase de linfocitos "deam-

buladores" que controlan constantemente las células y los tejidos de nuestro organismo. Si estos linfocitos encuentran una célula defectuosa, dan la voz de alarma en los ganglios linfáticos, poniendo así en guardia a una clase de linfocitos conocidos como células *T* "matadoras", que atacan y destruyen las células defectuosas. Entonces, las células *T* "matadoras" destruyen estas células de forma muy semejante o idéntica a como matan a células extrañas durante el rechazo de transplantes.

Sin embargo, hay algunas diferencias importantes entre el rechazo de un injerto y el mecanismo postulado de vigilancia inmunológica. En el rechazo de injertos, las dianas de las células *T* matadoras son claramente los antígenos *H*-2 del donante, que les son extraños, mientras que en el caso de las propias células anormales los antígenos *H*-2 de las células *T* y las células diana son idénticos. En el ataque a las células anormales, propias del huésped, los linfocitos *T* deben reconocer algún rasgo, verbigracia, un componente vírico (un antígeno viral) o un componente específico de una célula tumoral (un antígeno específico de tumor), que les permita diferenciar las células normales de las que se hallan alteradas.

Hasta hace poco no había pruebas directas de que los antígenos *H*-2 pudiesen desempeñar un papel en este proceso de reconocimiento. Entonces, Rolf Zinkernagel y Peter Dougherty, de la Australian National University, descubrieron que los linfocitos *T* matadores no podían destruir células infectadas por un virus, a no ser que la célula matadora y la célula diana tuvieran al menos un antígeno *H*-2 en común. John Schrader, Edelman y yo mismo hallamos que esto también era cierto para la acción de las células *T* matadoras sobre las células cancerosas. Cuando los anticuerpos que se unen a los antígenos *H*-2 compartidos por ambas células se añaden al medio de incubación, los linfocitos matadores se vuelven incapaces de destruir las células cancerosas, debido presumiblemente a que ya no pueden reconocer los antígenos *H*-2. Experimentos adicionales mostraron que son los antígenos *H*-2 presentes en la superficie de las células diana, y no los de las células *T* matadoras, los que tienen que estar bloqueados por los anticuerpos para prevenir la destrucción de las células diana.

Se han propuesto dos mecanismos generales para explicar el papel de los antígenos *H*-2 en la destrucción de las células anormales por los linfocitos *T* matadores. Un grupo de hipótesis sugiere que los linfocitos *T* matadores son portadores de dos tipos de moléculas o receptores; uno que se une a los antígenos *H*-2 y otro que se une al antígeno anormal (sea éste el viral o el tumoral). Según esta hipótesis, es necesaria la fijación de ambos receptores para que la acción del linfocito pueda llevarse a término. Otro grupo de hipótesis sugiere que la célula matadora tiene tan sólo un tipo de receptor, si bien éste se acopla selectivamente a un antígeno híbrido que consiste en un antígeno *H*-2 unido a un antígeno anormal. Desde el punto de vista de la teoría, cabe una estrecha interacción de este tipo entre dos antígenos, ya que se sabe que la membrana celular no es una estructura rígida, sino semifluida, con antígenos y otras proteínas que recuerdan a bloques de hielo flotando en un mar de lípidos de la membrana. Los antígenos *H*-2 y los anormales se encuentran de este modo en un movimiento constante sobre la superficie celular, lo que permitiría fácilmente que entrasen en contacto a través de interacciones realizadas al azar.

Una forma de determinación de si los antígenos *H*-2 y los antígenos anormales se encuentran realmente unidos entre sí y se asocian específicamente sobre la superficie de las células consiste en realizar lo que se conoce como la prueba del "gorro" *(cap)*.

Si se añaden anticuerpos dirigidos contra moléculas de la superficie celular, como son los antígenos *H*-2, a una suspensión de células, dichos anticuerpos reaccionarán con las moléculas de antígeno, formando con ellas un agregado, a modo de remiendo, sobre la superficie celular. Todos los remiendos o parches pueden llegar a fundirse en un extremo de la célula y así constituir un casquete o "gorro". Los remiendos y los casquetes se detectan con la ayuda de anticuerpos marcados con un colorante fluorescente que brilla con un color específico o en el microscopio de fluorescencia. Fluorocromos de diferentes colores (por ejemplo, fluoresceína que es verde y rodamina que es roja) pueden emplearse para marcar anticuerpos de diferentes especificidades, así como para determinar la distribución de distintos tipos de antígenos en la misma célula.

Con el fin de dilucidar si dos antígenos diferentes se encuentran estrechamente asociados en la superficie celular, se induce primero la formación del casquete con un tipo de antígeno mediante la adición de anticuerpos específicos para éste, marcados con fluoresceína; luego se fijan las células para que no puedan aparecer otros casquetes. Seguidamente, se tratan las células con anticuerpos específicos para el segundo tipo de antígeno, habiéndose marcado estos segundos anticuerpos con rodamina. Si en el microscopio de fluorescencia la etiqueta roja de la rodamina y la verde de la fluoresceína se encuentran asociadas en el casquete, podemos concluir que los dos tipos de antígenos se hallan físicamente asociados y han formado el casquete juntos. Si, por el contrario, la etiqueta verde de la fluoresceína parece como un casquete y la roja de la rodamina se encuentra esparcida difusamente sobre la superficie de la célula, puede decirse que los antígenos son independientes.

Nosotros hemos realizado este tipo de prueba con antígenos *H*-2 y antígenos víricos presentes en la superficie de células tumorales que reaccionan con anticuerpos para el virus leucémico Rauscher. Los dos tipos de antígeno forman un casquete en el mismo lugar de la superficie de la célula, lo que sugiere que están estrechamente asociados. Aunque el experimento no es definitivo, ya que depende de la especificidad del anticuerpo involucrado, nosotros (y otros) hemos confirmado esta observación por otros métodos diferentes. Ello nos permite propugnar la hipótesis de que antígenos *H*-2 y antígenos extraños puedan formar estructuras híbridas, que se reconozcan preferentemente por células *T* matadoras.

Un corolario de esta hipótesis es que los antígenos *H*-2 poseen rasgos estructurales especiales, que les capacitan para interaccionar en la superficie celular con antígenos víricos o tumorales. En un terreno más especulativo, la variabilidad genética del sistema *H*-2 podría afectar en teoría a la interacción de antígenos *H*-2 con antígenos anormales y a su detección por células *T* matadoras.

Se sigue debatiendo todavía si el modelo de uno o el de dos receptores de reconocimiento de las células *T* matadoras es el correcto. En cualquier caso, nuestro trabajo seguirá centrándose en el papel desempeñado por los antígenos *H*-2 en los mecanismos de la vigilancia inmunológica.

Cualquiera que llegue a ser la función concreta de los antígenos de histocompatibilidad, dicha función será de gran importancia para la inmunología y puede tener profundas implicaciones prácticas para el transplante de tejidos y el control de las enfermedades infecciosas y malignas. La medicina y el análisis molecular han llegado a ser compañeros en la teoría y en la aplicación práctica.

La solución del problema del mapa de cuatro colores

Bastan cuatro colores para colorear un mapa planar, de modo que nunca dos países adyacentes sean de igual color. Se ha probado la veracidad de esta famosa conjetura mediante una demostración de nuevo cuño, basada en ordenadores

Kenneth Appel y Wolfgang Haken

En 1852, pocos meses después de terminar sus estudios en el University College of London, Francis Guthrie escribió a su hermano Frederick, todavía en el "College", y discípulo del matemático Augustus de Morgan. En su carta, Francis hacía notar a Frederick que, al parecer, todo mapa trazado sobre una hoja de papel podía colorearse, usando solamente cuatro tintas, de modo tal que los países con frontera común tuvieran diferente color. Francis preguntaba si sería posible demostrar matemáticamente este hecho. Frederick no lo sabía, y le preguntó a De Morgan, quien lo ignoraba también. Durante los 124 años transcurridos desde entonces, el problema de los cuatro colores de Guthrie, a saber, el problema de demostrar que ningún mapa requiere más de cuatro colores, o de diseñar un mapa que precise por lo menos cinco, vino intrigando a matemáticos profesionales y aficionados, y desafiando a todos los jóvenes estudiantes, frecuentemente convencidos de que la única razón de que existan problemas no resueltos es la incompetencia de la generación precedente.

Nosotros resolvimos el problema de los cuatro colores en 1976. La conjetura de Guthrie ha sido demostrada matemáticamente, pero en forma muy diferente a la que él pudiera haber esperado. Incluso hoy en día hay un sentimiento de desánimo y consternación entre aquellos matemáticos que no estuvieron al corriente de los desarrollos que condujeron a la demostración del resultado. Para solucionar el problema se recurrió, en forma que no tiene precedentes, a modernos ordenadores; la demostración requiere cómputos muchos más largos que los tradicionalmente considerados aceptables. De hecho, no puede comprobarse la veracidad de la demostración sin auxilio de un ordenador. Más todavía, ciertas ideas cruciales de la demostración se perfeccionaron mediante experimentos computarizados. Desde luego, puede ser que algún día se descubra una demostración breve del teorema de los cuatro colores, quién sabe si por uno de esos brillantes jóvenes universitarios. También es concebible que no exista ninguna demostración de ese tipo. En este caso ha aparecido una nueva e interesante categoría de teoremas, que carecen de demostración en el sentido tradicional.

A pesar de los aspectos más novedosos de su demostración, tanto el problema de los cuatro colores, como el método básicamente utilizado para demostrarlo, están profundamente enraizados en las matemáticas. Comenzaremos a examinarlos retornando a la formulación inicial del problema, tal como fue enunciado en la carta de Guthrie. Al decir "países vecinos", Guthrie quiso sin duda decir "adyacentes" a lo largo de una línea fronteriza común a ambos, y no sólo en contacto por un punto; de lo contrario, un mapa cuyos países convergieran en un punto, como las raciones de una tarta, exigiría tantos colores como países. Al hablar de "países", ciertamente quiso decir "regiones conexas", pues si se permite que un país esté formado por más de una región, y éstas se encuentran separadas, no es nada difícil construir un ejemplo de mapa de cinco países en el que cada uno de ellos sea adyacente a los otros cuatro *(véase la ilustración superior de la página 80)*.

Es completamente seguro que Guthrie y De Morgan se dieron cuenta de que puede dibujarse un mapa con cuatro países en el que cada país sea adyacente a los otros tres *(véase la figura de la izquierda de la ilustración inferior de la página 80)*. Tal mapa exige cuatro colores, y, por tanto, una posible conjetura de tres colores es falsa. No bastarán tres colores para entintar todos los mapas.

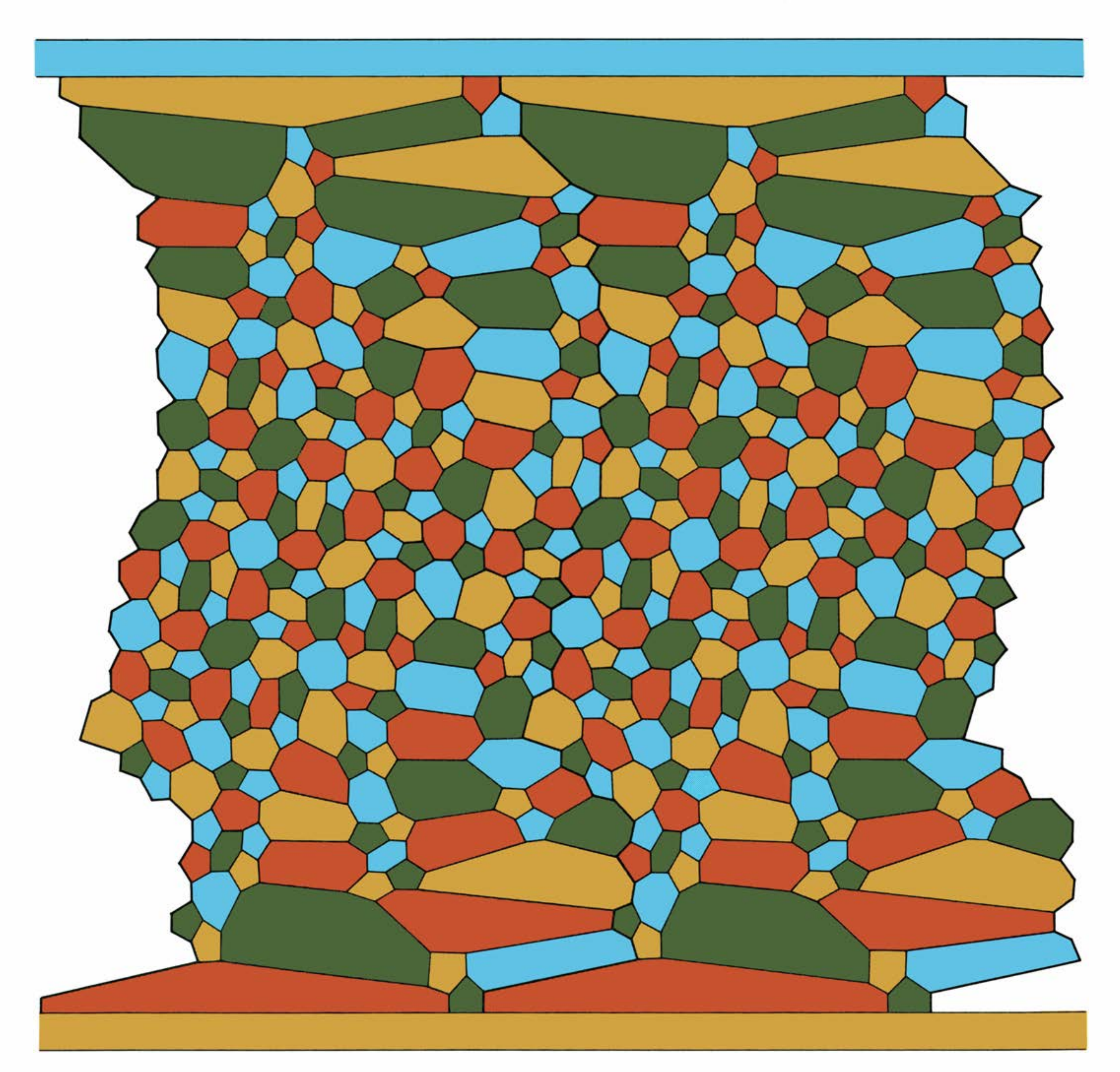
De Morgan demostró que no es posible que cinco países estén situados de forma tal que cada uno de ellos sea adyacente a los otros cuatro. Este resultado le hizo creer que nunca se necesitarían cinco colores, y que, así pues, la conjetura de los cuatro colores era verdadera. Demostrar que en un mapa no pueden existir cinco países mutuamente adyacentes no prueba, sin embargo, la conjetura de los cuatro colores *(véase la ilustración inferior de la página 80)*. Muchos matemáticos aficionados han descubierto por su cuenta demostraciones del resultado de De Morgan y creído haber demostrado así la conjetura de los cuatro colores, sin darse cuenta de este hecho.

En 1878, el matemático Arthur Cayley, viéndose incapaz de probar o refutar la conjetura de los cuatro colores, presentó el problema en la London Mathematical Society. Antes de pasar un año, Alfred Bray Kempe, abogado y miembro de la Sociedad, publicó un artículo en el que afirmaba haber demostrado que la conjetura era verdadera. El razonamiento de Kempe fue extraordinariamente lúcido, y a pesar de que su demostración resultó ser incompleta, contenía la mayor parte de las ideas fundamentales que un siglo después habrían de llevar a la demostración correcta. Kempe intentó demostrar la conjetura por el clásico método de *reductio ad absurdum;* supuso que la conjetura era falsa (esto es, que existía al menos un mapa que exigía cinco o más colores) y procedió a mostrar que tal hipótesis conducía a contradicción. De resultar una contradicción la hipótesis de partida (que algunos mapas requerían cinco colores) tenía que ser falsa, y en consecuencia, la conjetura de los cuatro colores (cuatro colores son siempre suficientes) era verdadera.

Kempe comenzó introduciendo la noción de mapa normal. Un mapa es normal si ninguno de sus países engloba

otros países y si, además, nunca hay más de tres países que tengan un punto común *(véase la ilustración de la página 81)*. A continuación, demostró que si hubiera un mapa que precisara cinco colores, un "mapa pentacromático", existiría también un mapa pentacromático normal. Así pues, para demostrar la conjetura de los cuatro colores es suficiente demostrar que no es posible obtener un mapa normal pentacromático. Kempe observó que si hubiera un mapa pentacromático normal entonces tendría que existir un mapa de esta clase de número mínimo de países, un "mapa pentacromático normal y minimal". (Dicho de otra forma, todo mapa con menor número de países que el pentacromático minimal podría colorearse con cuatro tintas o menos.) Así pues, para demostrar la conjetura de los cuatro colores es suficiente demostrar que un mapa normal, pentacromático y minimal, es imposible, esto es, que de postular la existencia de un mapa normal y minimal que exija cinco colores se deduce una contradicción.

Kempe intentó hallar la contradicción del modo siguiente. Demostró primero que todo mapa normal ha de tener un país con cinco, o menor número, de países vecinos. Después razonó que si un mapa normal pentacromático minimal tiene un país fronterizo con menos de seis vecinos (lo que, como acaba de probar, tiene que ocurrir en todo mapa normal), entonces habría de existir un mapa



MAPA DE 846 PAISES creado por Edward F. Moore, de la Universidad de Wisconsin, que se ha coloreado con cuatro tintas para explicar el teorema de los cuatro colores. Aunque algunos mapas (éste entre ellos) son francamente difíciles de colorear con sólo cuatro tintas, nadie ha creado jamás un mapa que exija forzosamente cinco o más colores. Hasta el año pasado, sin embargo, todos los intentos de que se tenga noticia, encaminados a demostrar que cuatro colores bastan para colorear cualquier mapa trazado sobre un plano o una esfera, habían fracasado. La dificultad de colorear un mapa depende del modo en que sus países sean fronterizos. Las "configuraciones", esto es, las disposiciones de países colindantes del mapa de Moore ayudaron a los autores a calibrar la magnitud de las dificultades de cómputo que aparejaría su método de demostración, que finalmente fue fructífero. En la ilustración sólo se muestra parte del mapa de Moore. El mapa completo es una proyección cilíndrica con octógonos en los polos norte y sur, y está formado por 54 octógonos, 228 heptágonos, 96 exágonos y 408 pentágonos.

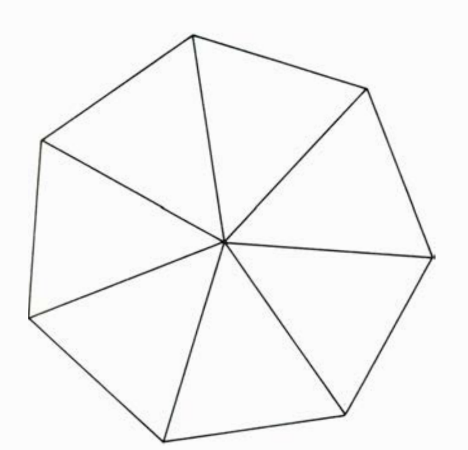
normal de menor número de países que también sería pentacromático. Si su razonamiento hubiera sido totalmente correcto se hubiera llegado a una contradicción. En la hipótesis de existencia de un mapa minimal pentacromático Kempe hubiera demostrado que existía un mapa pentacromático estrictamente menor, lo cual contradiría la definición de mapa minimal pentacromático, con lo que un mapa tal sería imposible. Como de ahí se deducía que no podría existir absolutamente ningún mapa pentacromático, la demostración de la conjetura estaría completa. Kempe demostró correctamente que si en un mapa pentacromático minimal existe un país que tenga dos, tres, o cuatro países fronterizos resulta una contradicción, pero su demostración para el caso de cinco países vecinos era defectuosa. En nuestra demostración del teorema de los cuatro colores corregimos las deficiencias del incorrecto análisis que Kempe dio del último caso, examinando para ello alrededor de 1500 disposiciones de los países. Fundamentalmente, nuestros métodos fueron extensión de partes válidas de la demostración de Kempe, que han sido objeto de gran atención y minucioso refinamiento por los matemáticos de los 100 últimos años.

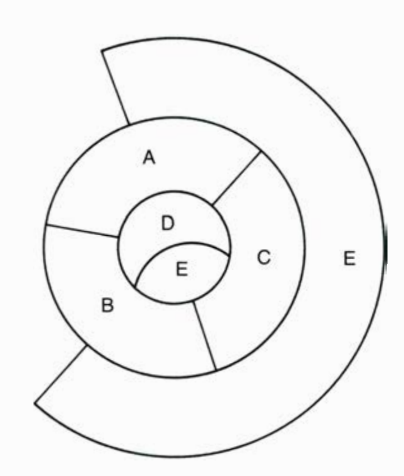
Kempe había demostrado que en todo mapa normal existe al menos un país que tiene dos, tres, cuatro o cinco vecinos. (Dicho de otra forma, no existe ningún mapa normal en el plano en el que todo país tenga seis o más países vecinos.) Lo anterior puede expresarse también diciendo que el conjunto de configuraciones formadas por un país con dos vecinos, por un país con tres vecinos, un país con cuatro vecinos, y un país con cinco vecinos *(véase la ilustración superior de la página 83)* es "inevitable", esto es, todo mapa normal ha de contener al menos una de estas cuatro configuraciones. La noción de configuración "inevitable" es una de las dos ideas fundamentales de nuestra demostración del teorema de los cuatro colores.

La segunda noción importante es la de configuración reducible. Una configuración es intuitivamente reducible si por puro examen de la configuración y del modo en que pueden alinearse cadenas de países se puede mostrar que la configuración no puede formar parte de ningún mapa minimal pentacromático. Kempe demostró que tres de sus cuatro configuraciones son reducibles, pero falló al querer demostrar que la configuración formada por un país con cinco vecinos era reducible. Los métodos diseñados para demostrar la reducibilidad de configuraciones han sido elaborados a partir de demostraciones que Kempe dio para la imposibilidad de mapas pentacromáticos minimales con países con cuatro fronterizos. El término "reducible" procede de la forma del razonamiento de Kempe: demostró que si un mapa pertacromático minimal contenía un país con, por ejemplo, cuatro vecinos, entonces podía reducirse el número de países del mapa sin que éste perdiera su carácter pentacromático *(véase la ilustración inferior de la página 83 y la ilustración de la página 84)*.

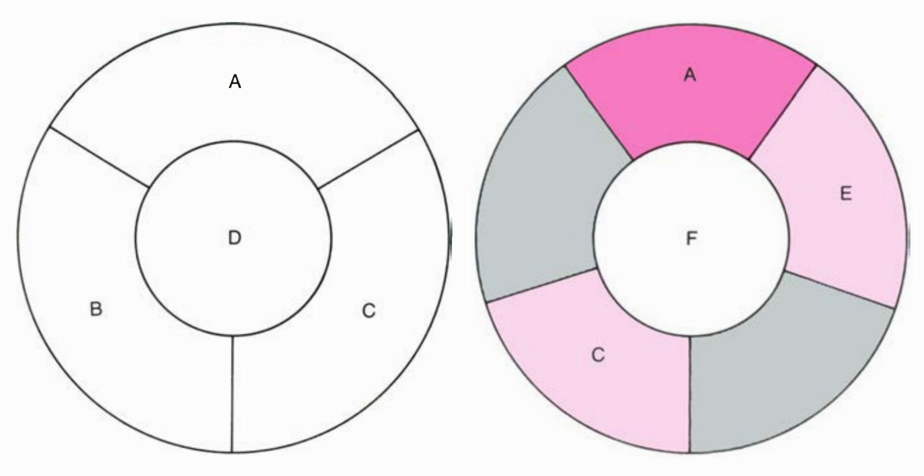
El método de Kempe para probar la conjetura de los cuatro colores puede describirse como el intento de hallar un conjunto inevitable de configuraciones reducibles. Sería suficiente hallar tal conjunto para tener demostrada la conjetura de los cuatro colores, porque ello probaría que todo mapa contiene una configuración que no puede formar parte de ningún mapa pentacromático minimal, y esto implica, como correctamente demostró Kempe, que no puede existir absolutamente ningún mapa pentacromático. Al igual que Kempe, nuestra forma de abordar el problema ha sido construir un conjunto inevitable de configuraciones reducibles. Sin embargo, en vez de las cuatro configuraciones sencillas de Kempe, nuestro conjunto está formado por unas 1500 configuraciones complejas.

En 1890, once años después de que Kempe publicara su demostración, Percy John Heawood señaló que el razonamiento de Kempe sobre la imposibilidad de mapas minimales pentacromáticos que contengan países con cinco vecinos era insuficiente, y que el error, además, no parecía de fácil arreglo. Analizando el problema, Heawood consideró otras generalizaciones de la conjetura de cuatro colores original. Los mapas estudiados





EL TEOREMA DE LOS CUATRO COLORES enuncia que bastan cuatro colores para iluminar cualquier mapa de forma que nunca haya dos países fronterizos del mismo color. Para que el teorema tenga sentido, los mapas deben estar formados por países en contacto. Los países vecinos deben ser adyacentes a lo largo de una línea, porque, si se considerasen como vecinos países que tuvieran solamente un punto de contacto, entonces el mapa de la izquierda precisaría colores distintos para cada uno de sus siete países. Debe entenderse que cada país es una sola región conexa, pues de lo contrario el mapa de la derecha, donde el país *E* consta de 2 piezas separadas, exigiría 5 colores.



TRES COLORES NO BASTAN para colorear cualquier mapa. Así lo demuestra el mapa de cuatro países dibujado a la izquierda. Por otra parte, tampoco es correcto suponer que el número de colores necesarios para un mapa es igual al máximo número de países mutuamente adyacentes del mapa. En el mapa de la derecha, nunca son mutuamente adyacentes más de tres países, a pesar de lo cual se necesitan cuatro colores: tres para el anillo exterior de países, y uno para el país central.

por Guthrie y Kempe eran mapas de un plano o de una esfera. Heawood, estudiando mapas trazados sobre superficies más complicadas, pudo obtener un elegante razonamiento que le permitió dar una cota superior para el número de colores necesarios para entintar mapas de estas superficies. Si el método que utilizó hubiera sido aplicable en el plano, se hubiera tenido una demostración de la conjetura de los cuatro colores.

Heawood siguió trabajando en el problema durante no menos de 60 años. En este lapso de tiempo, muchos otros eminentes matemáticos dedicaron al problema grandes esfuerzos. Puede uno preguntarse por qué tantos matemáticos invirtieron tanto tiempo en una cuestión de, aparentemente, tan magra importancia práctica. La explicación se encuentra en los descubrimientos que sobre la naturaleza de la matemática pura se realizaron a lo largo del siglo pasado y en el efecto que dichos descubrimientos tuvieron para reafirmar la fe de los matemáticos en la potencia de la matemática.

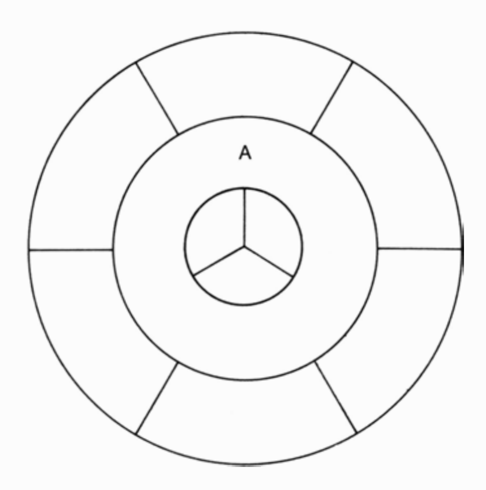
Hacia el fin del siglo XIX, los matemáticos habían podido construir poderosas teorías que permitieron elucidar numerosas cuestiones difíciles. Se fortaleció en ellos la creencia de que toda cuestión que pudiera formularse razonablemente en lenguaje matemático podría responderse utilizando ideas matemáticas de suficiente fuerza. Además, se creyó que estas cuestiones admitirían respuestas que un matemático competente podría comprobar por sí mismo en un tiempo razonable. Desde luego, el problema de los cuatro colores parecía ser una de tales cuestiones. El hecho de que los matemáticos no supieran resolverlo parecía claro que era debido a que todavía no habían puesto a punto el instrumental matemático adecuado.

Sin embargo, en la década de 1930 nuevos descubrimientos hicieron tambalearse la fe decimonónica en la completitud de las matemáticas. Kurt Gödel y Alonzo Church obtuvieron nuevos e inquietantes resultados en lógica formal, rama de las matemáticas donde el concepto de demostración matemática recibe su formulación más rigurosa. Se demostró que en el aparentemente más natural de los sistemas lógicos existen enunciados verdaderos cuya veracidad o falsedad no puede ser demostrada dentro del sistema. Además, el sistema contiene teoremas cuyos enunciados son relativamente breves, pero cuyas demostraciones son tan largas que sería imposible consignarlas en un período de tiempo razonable. Una investigación posterior, hacia la década de 1950, mostró que idénticas dificultades a las surgidas en la lógica formal aparecían en otras ramas de las matemáticas. Algunos matemáticos empezaron a considerar la idea de que la conjetura de los cuatro colores podría ser uno de esos enunciados tan imposibles de demostrar como de refutar; después de todo, la conjetura había sido estudiada infructuosamente durante 80 años. Otros especialistas opinaban que de existir una demostración, sería imposible presentarla por su enorme longitud. Muchos, sin embargo, siguieron creyendo que esta región de las matemáticas se encontraba a salvo de la enfermedad de indecibilidad, y que sería posible encontrar un elegante razonamiento matemático que decidiera la veracidad o falsedad de la conjetura.

Sabemos ahora que puede hallarse una demostración. Pero lo que todavía no sabemos (y quizá no podamos saber nunca) es si existe una demostración que sea elegante, concisa y completamente comprensible por una mente matemática humana. En los diversos intentos de demostrar la conjetura de los cuatro colores han intervenido tantas ramas de la matemática que sería imposible comentarlas todas aquí. Nos ceñiremos a exponer los trabajos que condujeron directamente a la demostración.

En 1913, George D. Birkhoff, de la Harvard University, mejoró la técnica de reducción de Kempe, y pudo demostrar que ciertas configuraciones, mayores que las de Kempe, eran reducibles. Utilizando algunos resultados de Birkhoff, Philip Franklin, del Instituto de Tecnología de Massachusetts, pudo demostrar que todo mapa que requiriese cinco colores habría de estar formado por 22 países como mínimo, es decir, que todo mapa de menos de 22 países es coloreable con sólo cuatro tintas. Los métodos de Birkhoff fueron mejorados entre 1913 y 1950 por muchos matemáticos. Durante este período la noción de configuración reducible se aprovechó, sobre todo, para mejorar el resultado de Franklin; hacia 1950 se había demostrado que todo mapa de menos de 36 países es tetracoloreable.

A pesar de que el trabajo realizado sobre esta materia demostró que muchas configuraciones son reducibles, el conjunto de todas las configuraciones cuyo carácter reducible estuviera bien establecido no llegaba en 1950 ni de lejos a formar un conjunto inevitable. Tan sólo unos cuantos matemáticos habían construido conjuntos de configuraciones inevitables, y no se tenían grandes esperanzas de que su trabajo pudiera conducir a un conjunto inevitable de configuraciones reducibles.



LOS MAPAS "NORMALES" fueron definidos por Alfred Bray Kempe, quien publicó en 1879 una demostración insuficiente del teorema de los cuatro colores. Un mapa es normal si ninguno de sus países rodea completamente a otro u otros países del mapa, y, además, nunca hay más de tres países que concurran en un punto. El mapa de esta figura no es normal, porque el país *A* encierra a otros tres países. Kempe probó correctamente que si se pudiera demostrar el teorema de los cuatro colores para mapas normales, el teorema sería cierto para mapas cualesquiera. Los autores consideraron mapas normales en su demostración del teorema.

Heinrich Heesch, de la Universidad de Hannover, parece haber sido el primer matemático que con posterioridad a Kempe afirmó públicamente que la conjetura de los cuatro colores podría demostrarse hallando un conjunto inevitable de configuraciones reducibles. Heesch comenzó a trabajar en la conjetura en 1936, y ha contribuido con varias aportaciones fundamentales al actual estado de la teoría. En 1950 estimó que las configuraciones reducibles del (entonces) hipotético conjunto inevitable sería de ciertos tamaños, y no arbitrariamente grandes, y que su número sería del orden de 10.000.

El tamaño de las configuraciones es, necesariamente, una de las principales cuestiones a considerar al atacar el problema de los cuatro colores por este procedimiento. En el siglo transcurrido desde que Kempe introdujo por vez primera la noción de reductibilidad se han desarrollado ciertos métodos estándar para examinar configuraciones al objeto de determinar si son reducibles o no. Al aplicar estos métodos a configuraciones grandes es preciso comprobar gran número de detalles, lo que solamente parece posible con auxilio de un ordenador. Toda configuración está rodeada por un anillo de países vecinos. El tamaño del anillo, es decir, el número de países que componen el anillo que rodea la configuración, guarda relación directa con la dificultad para demostrar que la con-

figuración es reducible. Cuando se está tratando de construir un conjunto inevitable de configuraciones reducibles, y se descubre que una cierta configuración es irreducible, suele ser posible sustituirla con buenos resultados por una o más configuraciones distintas, configuraciones que, por lo general, tendrán un anillo de tamaño mayor. Sin embargo, al sustituir una configuración por otra cuyo anillo contenga un vértice más aumenta considerablemente la dificultad de la comprobación de reductibilidad, así como el tiempo de ordenador necesario. Ello es debido, en parte, a que el número de coloreados correctos de los vértices del nuevo anillo es aproximadamente tres veces mayor que el número de posibles coloreados de los vértices del anillo primitivo. Además, los programas de computador preparados al efecto de comprobar la reductibilidad de una configuración toman en cuenta varias veces cada uno de los posibles métodos de coloreado. En 1950, las dificultades de cómputo parecían descartar la posibilidad de obtener un conjunto inevitable de configuraciones y poder demostrar que cada uno de sus componentes era reducible.

La aparición de los ordenadores digitales de alta velocidad, sin embargo, hizo técnicamente posible abordar tales problemas. Heesch formalizó los métodos hasta entonces conocidos para demostrar la reductibilidad de configuraciones, y observó que al menos uno de ellos (una generalización directa del método empleado por Kempe) era en principio lo suficientemente mecánico como para poder encomendarlo a un ordenador. Un discípulo de Heesch, Karl Dürre, preparó entonces un programa de ordenador que aplicaba dicho principio como criterio de reductibilidad. Cuando este programa comprueba que una configuración es reducible de acuerdo con el criterio, la configuración es verdaderamente reducible. Pero un resultado negativo tan sólo muestra que este criterio particular de verificación de reductibilidad no es suficiente para decidir si la configuración es reducible o no; tal vez pudiera demostrarse que es reducible utilizando otros métodos. En algunas ocasiones en que el programa de Dürre fracasó en demostrar que una configuración era reducible, Heesch tuvo éxito. Pudo demostrar que dichas configuraciones eran reducibles mediante datos generados por el programa junto con algunos cálculos adicionales, empleando una técnica más potente preparada por Birkhoff.

Heesch utilizó un método muy conveniente para la descripción de configuraciones. Comenzó por transformar el mapa original en lo que los matemáticos llaman su grafo dual: un grafo planar donde cada vértice del grafo representa un país, y cada trazo entre dos vértices representa una frontera. Para obtener el grafo dual de un mapa se señala la capital de cada país, y después, siempre que dos países sean vecinos, se unen sus capitales por una carretera que atraviese la frontera común *(véase la ilustración de la página 85)*. A continuación se suprime todo, a excepción de las capitales (llamadas vértices) y de las carreteras (llamadas lados), recién añadidas. Estos vértices y lados forman el grafo dual del mapa original. Los lados de un grafo dividen al plano en regiones llamadas caras. Si el mapa original es normal (y para el teorema de los cuatro colores basta estudiar mapas normales) todas las caras del grafo dual son triángulos. En este caso se dice que el grafo dual es una triangulación. El número de lados que concurren en un vértice dado (del grafo dual) se llama grado del vértice, y es igual al número de países vecinos que en el mapa original tenía el país correspondiente al vértice. Un camino (circuito) formado por lados que partiendo de un vértice termine en ese mismo vértice sin cortarse a sí mismo descompone al grafo en dos partes: su exterior y su interior.



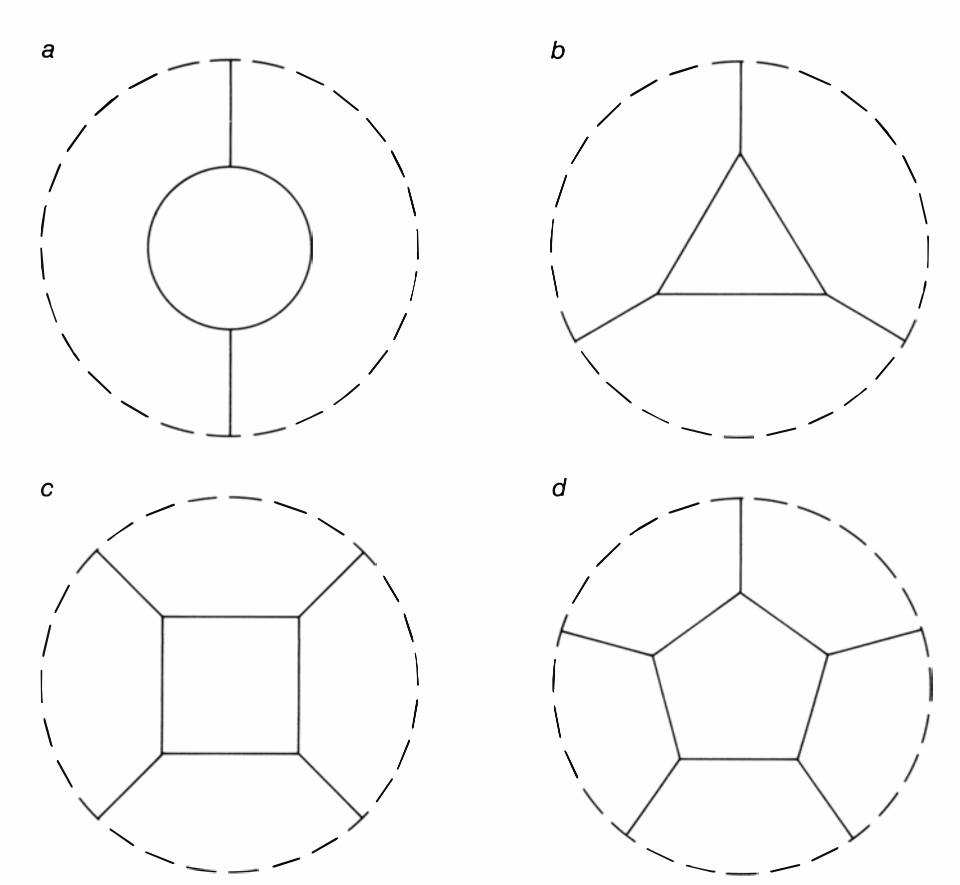
EL MAPA DE LOS ESTADOS ORIENTALES de los Estados Unidos es normal, pero el mapa total del territorio continental estadounidense no lo es: Utah, Colorado, Arizona y New México se reúnen en un solo punto. (Obsérvese que los 48 estados continentales contiguos ni siquiera constituyen un mapa en sentido estricto, porque el estado septentrional de Michigan está formado por dos piezas que se encuentran desconectadas.)

Con el vocabulario de grafos duales, una configuración es una parte de una triangulación formada por un conjunto de vértices más todos los lados que los conectan. El circuito frontera (perimétrico) formado por todos los vértices adyacentes a la configuración, juntamente con todos los lados que los unen, se llama anillo de la configuración. (El anillo del grafo corresponde al anillo de países que limitan la configuración en el mapa primitivo.) Las configuraciones suelen describirse por las longitudes de sus anillos: por ejemplo, una configuración se llama de anillo-seis cuando su circuito frontera exactamente consta de seis vértices.

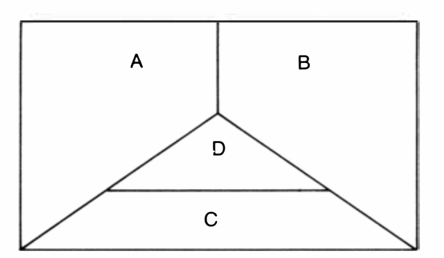
Con los trabajos de Heesch la teoría de configuraciones reducibles parecía hallarse extremadamente bien desarrollada. A pesar de haberse introducido algunos perfeccionamientos en los métodos de verificación de configuraciones reducibles, todas las ideas relativas a la noción de reductibilidad que se necesitan en la demostración del teorema de los cuatro colores estaban claramente comprendidas a finales de la década de 1960. En cambio, no se habían realizado progresos comparables en lo referente al descubrimiento de conjuntos inevitables de configuraciones. Con el objeto de hallar conjuntos inevitables de configuraciones (no necesariamente reducibles), Heesch introdujo un nuevo método, que puede asimilarse al desplazamiento de cargas eléctricas en una malla; pero la noción de conjunto inevitable no suscitó en él el mismo entusiasmo que las configuraciones reducibles. Ello no obstante, el método de "descarga", que apareció por vez primera en forma bastante rudimentaria en los trabajos de Heesch, ha resultado crucial para todo lo posteriormente realizado sobre conjuntos inevitables. Aunque en forma mucho más compleja, llegó a convertirse en el elemento central de la demostración del teorema de los cuatro colores; por esta razón lo expondremos con cierto detalle.

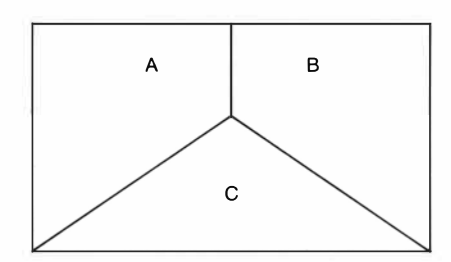
El trabajo de Kempe muestra que una triangulación correspondiente a un mapa minimal pentacromático no puede contener vértices con menos de cinco vecinos. En lo que sigue, reservamos por comodidad el término triangulación para las triangulaciones sin vértices de grado menor que cinco. Asignándole a cada vértice de grado *k* (esto es, con *k* vecinos) el índice de carga 6-*k*, los vértices de grado mayor que seis (llamados vértices mayores) tendrán asignada carga negativa, y solamente los vértices de grado cinco tendrán asignada carga positiva. Del trabajo de Kempe se deduce que la suma de los números índice de todas las triangulaciones es exactamente igual a 12. Este sorprendente resultado es consecuencia de ser el grafo una triangulación en el plano. No tiene gran importancia que la suma sea exactamente doce. Lo que sí es extremadamente importante es que la suma de carga sea estrictamente positiva para toda triangulación planar.

Supongamos ahora que las cargas de tal triangulación se redistribuyan y desplacen, sin que la carga total del sistema sufra alteración. Más concretamente, supongamos que se desplacen cargas positivas desde algunos vértices cargados

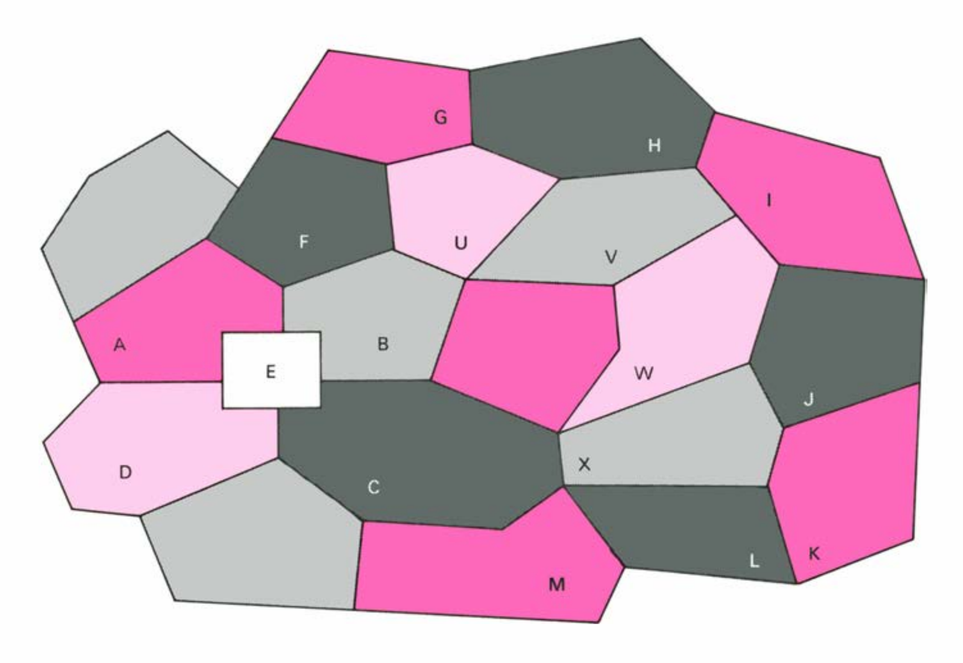


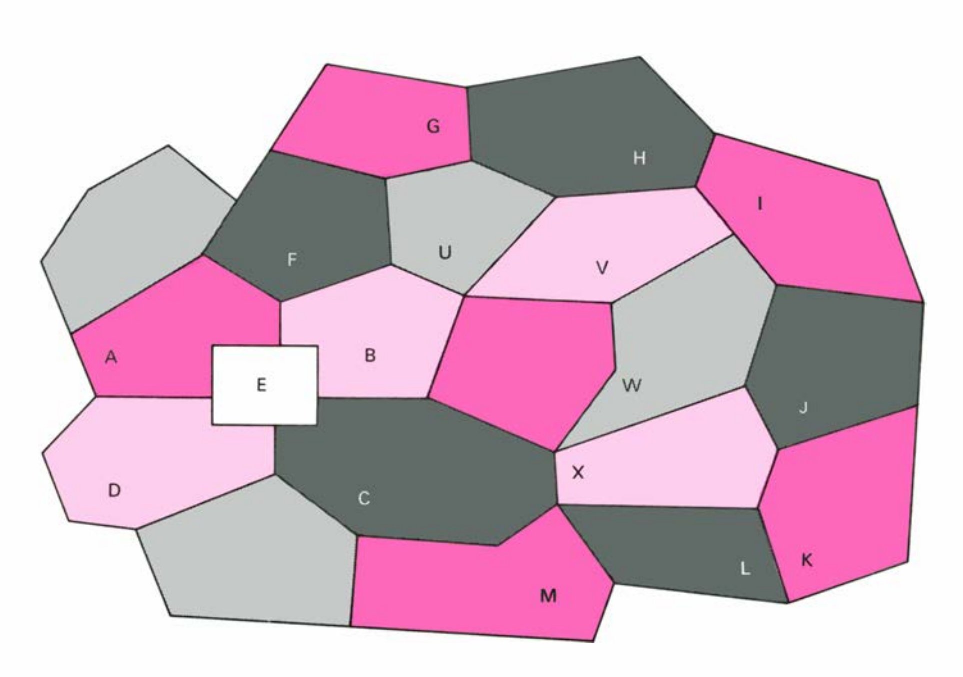
CONJUNTO DE CUATRO CONFIGURACIONES que, según demostró Kempe, es "inevitable"; es decir, todo mapa normal contiene al menos uno de los miembros de este conjunto. Las configuraciones del conjunto son un país con dos vecinos *(a)*, un país con tres vecinos *(b)*, un país con cuatro vecinos *(c)*, y un país con cinco vecinos *(d)*. Como es obvio, ningún país puede tener cero vecinos o sólo un vecino, porque en un mapa normal no puede haber islas ni países rodeados por otro.





KEMPE DEMOSTRO que, para probar el teorema de los cuatro colores, es suficiente demostrar que no puede existir ningún mapa minimal pentacromático. Un mapa minimal pentacromático sería un mapa que exigiendo cinco colores tuviera mínimo número de países, con lo que todo mapa menor (con menos países) podría colorearse con cuatro colores o menos. Kempe intentó demostrar que los mapas minimales pentacromáticos son imposibles probando para ello que ninguna configuración de su conjunto inevitable puede formar parte de un mapa minimal pentacromático. Todo mapa planar contiene alguna de estas configuraciones, por lo que si su razonamiento hubiese sido correcto habría demostrado el teorema. He aquí la demostración de imposibilidad de que un país, *D*, con tres vecinos *(izquierda)* no puede formar parte de un mapa minimal pentacromático, tal como fue correctamente concebida por Kempe. Se combina *D* con uno de sus vecinos, *C*, y se forma un nuevo país, *C' (a la derecha)*. El nuevo mapa tiene menos países que el mapa minimal pentacromático de partida, y, por tanto, puede colorearse con cuatro colores. Ahora se les da a todos los países, exceptuado *D*, los colores que les corresponderían usando cuatro tintas para el mapa simplificado. *D* puede entonces colorearse con la tinta no empleada para *A*, *B* o *C*. Por consiguiente, el mapa primitivo puede iluminarse con cuatro colores: no es un mapa minimal pentacromático. Un razonamiento parecido muestra que ningún país con dos vecinos puede ser parte de un mapa minimal pentacromático. La demostración para el país con cuatro vecinos se presenta en la ilustración siguiente.





UNA CONFIGURACION REDUCIBLE es una disposición de países que no puede ser parte de ningún mapa minimal pentacromático. Para demostrar el teorema de los cuatro colores Kempe intentó probar que las cuatro configuraciones de su conjunto inevitable eran reducibles. Fracasó porque no pudo demostrar que la configuración formada por un país con cinco vecinos es reducible. Los autores consiguieron demostrar el teorema construyendo un conjunto inevitable formado por alrededor de 1500 configuraciones reducibles. La noción de reductibilidad se extrajo de la demostración, correctamente concebida por Kempe, de que un país con cuatro vecinos, como por ejemplo el *E*, no puede formar parte de un mapa minimal pentacromático. Al igual que en el caso de un país con tres vecinos *(véase la ilustración inferior de la página anterior)* combinando *E* con un país vecino se puede establecer una coloración de cuatro tintas para los otros países del mapa *(arriba)*. Si los cuatro vecinos de *E* reciben menos de tres colores, se puede asignar a *E* el color no utilizado, obteniéndose una coloración de cuatro colores del mapa completo. En caso contrario, examinemos los colores de un par de países en lados opuestos de *E*. O bien existe un camino formado por países teñidos con estos dos colores que conecta los países del par considerado, o bien no existe tal camino. En nuestro ejemplo, *A* y *C* están conectados por un rosario de países de color rojo oscuro y gris oscuro; en cambio no hay ningún camino de colores claros que conecta *B* con *D*. Un teorema enuncia que si ambos pares de países están unidos por caminos de este tipo, entonces los caminos han de tener un país en común. Pero esto es obviamente imposible, porque el país común tendría que recibir dos coloraciones distintas. Así que al menos un par de países (aquí son *B* y *D*) no estarán enlazados por un camino de países con los colores de este par. Los países de color claro que forman un camino partiendo de *B* son *U, V, W* y *X*. Inviértase los colores de los países de este grupo. En este caso, hay que colorear de color claro los países gris claro, y de color gris claro los de rojo claro *(abajo)*. Ahora el país no coloreado está rodeado por vecinos de tres colores solamente, y puede recibir el cuarto color *(gris claro)* con lo que se obtiene una tetracoloración del mapa minimal pentacromático: un mapa con cuatro vecinos no puede ser parte de un mapa minimal pentacromático.

positivamente (de grado cinco) hasta algunos vértices cargados negativamente (vértices mayores). Ciertamente, no es posible cambiar la suma (positiva) de las cargas realizando estas operaciones, pero sí podrán cambiar los vértices con carga positiva; por ejemplo, algunos vértices de grado cinco pueden perder toda su carga positiva (descargarse), mientras que otros vértices mayores pueden recibir tanta carga que terminen con carga positiva (sobrecargarse.) Según el proceso de carga o descarga que se elija, distintos vértices quedarán descargados o sobrecargados.

Sin embargo, si para un grafo arbitrario se da un proceso de descarga específico, es posible hacer una lista finita de todas las configuraciones que, tras efectuar la descarga, contienen vértices con carga positiva. (Desde luego, cada una de estas configuraciones puede repetirse un número indefinido de veces.) Expresado de otra forma, solamente pueden presentar carga positiva las configuraciones pertenecientes a este conjunto finito. Como la carga total es siempre positiva, habrá siempre algunos vértices con carga positiva. Así pues, como todos los posibles receptáculos de carga positiva están incluidos en la lista de configuraciones anteriores, toda triangulación planar ha de contener al menos una de estas configuraciones. Este proceso –consistente en generar una lista de configuraciones específicas a partir de un mapa arbitrario– funciona bien porque es posible determinar la colocación de pequeños pedazos del mapa sin conocer la forma del mapa completo.

Para la demostración del teorema de los cuatro colores, lo que se persigue con la descarga de vértices positivos es hallar el sistema de describir exactamente los procesos de transporte de cargas que aseguren que todo vértice cargado positivamente en la configuración resultante tenga que pertenecer a una configuración reducible o que ser adyacente a una configuración reducible. Como las configuraciones señaladas por este procedimiento deben formar un conjunto inevitable, si también fueran reducibles la conjetura de los cuatro colores estaría demostrada. Desde luego, si alguna de las configuraciones resultantes no es reducible no se habría avanzado realmente nada. De hecho, el conjunto inevitable de Kempe puede considerarse obtenido por el ineficaz procedimiento de no desplazar carga ninguna.

Para aclarar las consideraciones anteriores expondremos un ejemplo bastante sencillo de proceso de descarga y de conjunto inevitable asociado. Consideremos

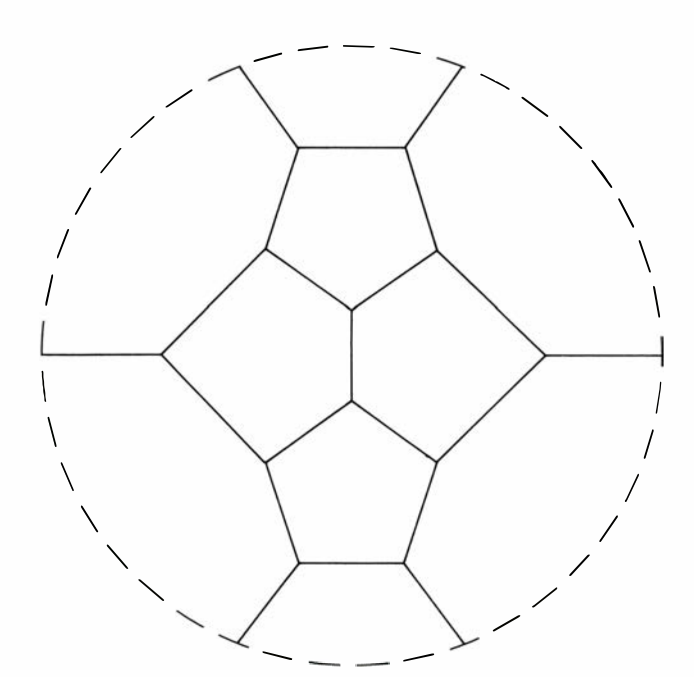
el proceso de descarga que transfiere 1/5 de unidad de carga desde cada vértice de grado cinco hasta cada uno de sus vértices mayores adyacentes *(véase la ilustración de la página 88)*. El conjunto inevitable asociado está formado por dos configuraciones. Una de ellas es un par de vértices de grado cinco unidos por un lado, y la otra es un vértice de grado cinco unido mediante un lado a un vértice de grado seis.

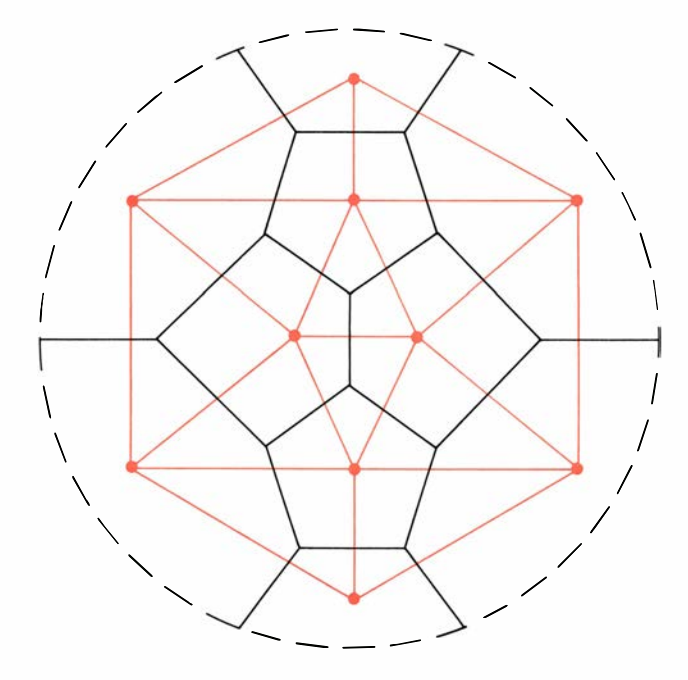
Estas configuraciones se han obtenido de la forma siguiente. La única forma de que un vértice de grado cinco termine con carga positiva en este procedimiento es que al menos uno de sus vecinos no sea un vértice mayor, con lo que el vértice se verá obligado a retener carga positiva; o bien el vértice tiene un vecino de grado cinco (situación correspondiente a la primera configuración del conjunto) o bien tiene un vecino de grado seis (segunda configuración).

Un vértice de grado seis comienza sin carga, y por consiguiente no puede recibir ninguna. La única forma de que un vértice de grado siete llegue a ser positivo es que tenga al menos seis vecinos de vértices de grado cinco; si tiene seis de tales vecinos (de grado cinco) al menos dos de ellos estarán unidos por un lado (primera configuración del conjunto inevitable). Un vértice de grado ocho o superior no puede llegar a ser positivo, aún cuando todos sus vecinos sean vértices de grado cinco. Para verlo examinemos qué sucede en los vértices de grado ocho. Todos tienen carga inicial igual a menos dos, y la máxima carga positiva que pueden recibir es ocho veces 1/5, o sea, 1 $^3/_5$. Así que las dos configuraciones (no reducibles) anteriores forman un conjunto inevitable; es decir, como estos cálculos son válidos en toda triangulación planar (sin vértices de grado menor que cinco), toda triangulación del plano (con la condición anterior) contendrá al menos un miembro del conjunto de dos configuraciones analizado.

Uno de nosotros dos (Haken) observó, en 1970, que existían métodos para perfeccionar los procesos de descarga, y comenzó a pensar que las mejoras introducidas podrían conducir a una demostración de la conjetura de los cuatro colores. Las dificultades, no obstante, parecían todavía formidables. Entre ellas se encontraba la creencia de que todo conjunto inevitable contendría algunas configuraciones muy grandes (con anillos de hasta 18 vértices vecinos); ello querría decir que el problema muy bien pudiera encontrarse por encima de las posibilidades de los ordenadores entonces existentes, pues aunque la verificación del carácter reducible de configuraciones reducibles de poco tamaño (hasta unos 11 vértices) era razonablemente sencilla de realizar mediante ordenador, el tiempo de cómputo empleado se multiplicaba por un factor de cuatro cada vez que el tamaño del anillo aumentaba en una unidad. Para empeorar las cosas, la capacidad de almacenamiento del ordenador tendría que aumentar a casi igual ritmo. Al aplicar el programa de Dürre a una configuración especialmente difícil de anillo-14, se precisaron 26 horas para demostrar que la configuración no satisfacía ni siquiera la más mecánica de las definiciones de reductibilidad (definición que se enuncia con el mínimo número de instrucciones a la máquina). Incluso si el tiempo promedio para examinar una configuración de anillo-14 fuera de solamente 25 minutos, la comprobación de una configuración de anillo-18 exigiría, en promedio, $4^4 \times 25$ minutos, o sea, más de 100 horas, de tiempo de ordenador, y mayor capacidad de almacenamiento de la disponible en ninguno de los ordenadores existentes.

Otra de las dificultades fundamentales consistía en que nadie sabía cuántas configuraciones reducibles serían precisas para disponer de un conjunto inevitable. Parecía verosímil que el número de configuraciones fuese del orden de varios millares, pero no se había podido establecer un límite superior. Supongamos que con un ordenador de suficiente capacidad sean precisas 100 horas para demostrar que una configuración de anillo-18 es reducible. Si el conjunto inevitable tuviera 1000 configuraciones de anillo-18 serían necesarias 100.000 horas, más de 11 años, de funcionamiento de un ordenador gigantesco para demostrar que todas ellas son reducibles. A efectos prácticos, si el conjunto hubiera sido de ese tamaño, la demostración hubiera





UN GRAFO DUAL *(en color)* **representa los países y las líneas fronterizas del mapa primitivo** *(izquierda)*. **El grafo dual se construye marcando un vértice en cada país del mapa primitivo y uniendo los vértices correspondientes a pares de países vecinos mediante un lado que atraviese su frontera común. Los lados pueden siempre trazarse con líneas rectas, que dividen al plano en caras poligonales. Cuando el mapa primitivo es normal las caras son triángulos, y el grafo se llama una triangulación del plano. El número de lados que se reúnen en un vértice común se llama grado del vértice, y es igual al número de países adyacentes (en el mapa original) que tenía el país que quede representado por el vértice.**

tenido que esperar a que existieran ordenadores mucho más veloces que los entonces disponibles.

Aun en el caso de que el teorema pudiera demostrarse hallando un conjunto inevitable de configuraciones reducibles, la demostración no sería satisfactoria para quienes exigen elegancia matemática. Y lo que sería todavía más molesto para muchos matemáticos: nadie podría comprobar personalmente la reductibilidad de todas las configuraciones del conjunto. No obstante, hacia 1970 la mayoría de los expertos en el problema de los cuatro colores eran francamente pesimistas acerca de la posibilidad de encontrar una demostración breve. El problema había recibido ya mucha atención desde que fuera formulado, más de cien años antes. Se habían ensayado ya muchas posibles vías de penetración, y aunque algunas de ellas condujeron a importantes resultados en otras ramas de las matemáticas, ninguna proporcionó jamás una demostración del teorema de los cuatro colores.

Cuando comenzamos, en 1972, nuestro trabajo en el problema estábamos convencidos de que las técnicas de que disponíamos no nos llevarían a una demostración no computarizada. Más aún, teníamos serias dudas de que dichas técnicas pudieran conducirnos a ningún tipo de demostración antes de que se desarrollasen ordenadores mucho más potentes. Así que el primer paso para abordar el problema de hallar un conjunto inevitable de configuraciones reducibles fue determinar si había o no esperanza de hallar tal conjunto con configuraciones cuyo tamaño anular fuera lo bastante pequeño como para que el tiempo de ordenador requerido en la verificación de reductibilidad se pusiera en razón. Por la naturaleza misma de la cuestión era evidente que no se debería comenzar examinando la reductibilidad de todas las configuraciones consideradas; de lo contrario, el tiempo requerido para efectuar la estimación sería mayor que el tiempo esperado necesario para ejecutar la tarea completa.

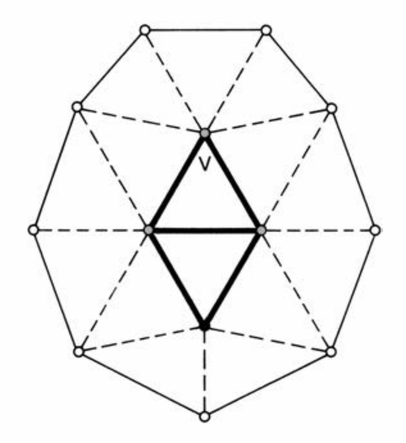
En este punto resultó ser de extraordinaria utilidad una idea de Heesch. Al efectuar la comprobación de reductibilidad de configuraciones, Heesch observó cierto número de fenómenos característicos que dan indicación de la verosimilitud de reducción fructífera. Por ejemplo, hay ciertas condiciones, en las que intervienen los vecinos de los vértices de la configuración, bajo las cuales nunca se ha encontrado una configuración reducible. Así, no se ha encontrado jamás una configuración reducible que contuviese al menos dos vértices, que poseyera un vértice adyacente a cuatro vértices del anillo, y que careciera de configuraciones reducibles más pequeñas. Aun cuando no se tiene demostración ninguna de que no puedan existir configuraciones que a pesar de estos obstáculos sean reducibles, parecía prudente suponer que si se querían configuraciones reducibles se deberían evitar tales configuraciones. Heesch descubrió tres importantes impedimentos cuya descripción era sencilla, entre los que se cuenta el anterior, que obstaculizan la reducción de configuraciones *(véase la ilustración de la página opuesta).* No se ha podido nunca demostrar que ninguna configuración que contenga alguno de estos obstáculos sea reducible.

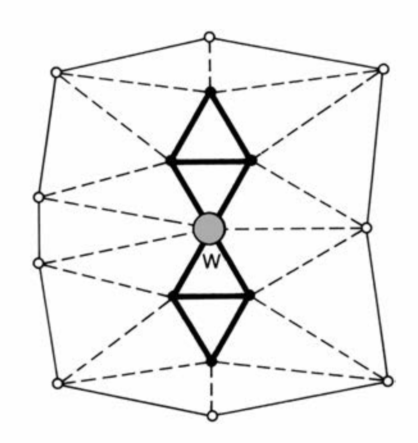
Es fácil determinar si una configuración contiene o no obstáculos de reducción de los tipos anteriores, y las configuraciones que carecen de ellos tienen buena probabilidad de resultar reducibles. Teníamos la impresión de que si hubiera un conjunto inevitable de configuraciones sin obstáculos de reducción que resultase manejable, habría un conjunto inevitable de tamaño parecido que contendría tan sólo configuraciones reducibles.

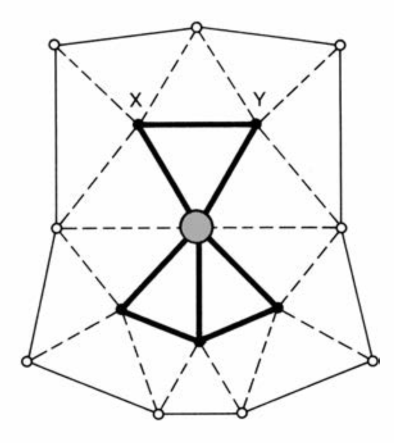
En consecuencia, decidimos estudiar ante todo ciertas clases de procesos de descarga, al objeto de determinar los tipos de conjuntos de configuraciones libres de obstáculos que pudieran surgir. Para mejorar nuestra comprensión de lo que nos hacía falta, restringimos, incluso para estudio, el problema ya restringido a las llamadas configuraciones geográficamente buenas: configuraciones que no contienen los dos primeros de los tres obstáculos descritos por Heesch.

En el otoño de 1972, escribimos un programa de ordenador que realizara el tipo especial de proceso de descarga que nos parecía más razonable. No estábamos todavía preparados para demostrar el teorema, así que la salida de nuestro programa no era un conjunto inevitable, sino más bien una lista de las configuraciones que aparecían en los casos más importantes. Aunque no puede esperarse de un programa de ordenador que actúe con la misma destreza que un ser humano, la inmensa velocidad del ordenador

UN EJEMPLO DE TRIANGULACION PLANAR que contiene un vértice de grado ocho *(cuadrado gris)*, un vértice de grado siete *(círculo gris grande)*, ocho vértices de grado seis *(círculos grises pequeños)* y 15 vértices de grado cinco *(círculos negros pequeños)*. Kempe demostró que los vértices de grados dos, tres y cuatro son reducibles, esto es, no pueden formar parte de un mapa minimal pentacromático. Por consiguiente, para demostrar la inexistencia de mapas minimales pentacromáticos los autores tan sólo tuvieron que tomar en consideración las triangulaciones con vértices de grado cinco o mayor. La "carga" *(cifras de color)* de un vértice es, por definición, igual a seis menos el grado del vértice. Como las triangulaciones examinadas no tienen vértices de grado cuatro o menor, los únicos vértices de carga positiva son los de grado cinco. No es difícil demostrar que la carga total de todo mapa normal es igual a 12. De aquí resulta que habrá vértices cargados positivamente en la triangulación del plano asociada a la demostración del teorema de los cuatro colores.







TRES OBSTACULOS DE REDUCCION observados por Heinrich Heesch, de la Universidad de Hannover. Estas disposiciones de los vértices parecen figurar solamente en configuraciones cuya reductibilidad no es posible demostrar. Por ello los autores se sirvieron de estos obstáculos para reconocer posibles zonas problemáticas de su demostración. En los grafos que aquí se muestran, las configuraciones son las disposiciones de vértices unidos por línea continua gruesa. Las líneas continuas finas conectan los vértices del anillo de la configuración, o sea, de su frontera externa. Las líneas de puntos conectan los vértices de la configuración con los vértices del anillo. Cada una de estas configuraciones contiene uno de los obstáculos de reducción de Heesch: el vértice *V* *(a la izquierda)* tiene cuatro vecinos en el anillo de la configuración; el vértice *W (centro)* tiene tres vecinos no consecutivos en el anillo de la configuración; los vértices *X* e *Y (a la derecha)* forman el llamado "par colgante"; es decir, estos vértices están unidos entre sí y a los vértices del anillo, pero sólo tienen un único vecino más en la configuración.

permite transigir en ciertas ineficiencias. Por si acaso, el programa se preparó de forma que los resultados pudieran comprobarse a mano sin dificultad.

Las primeras pasadas del programa, a finales de 1972, nos proporcionaron abundante y valiosa información. Se vio, en primer lugar, que cerca de la mayoría de los vértices con carga terminal positiva aparecían configuraciones geográficamente buenas de tamaño razonable, con anillos de no más de 16 países. En segundo lugar, se vio que se repetían con suficiente frecuencia las mismas configuraciones como para hacer pensar que la lista de configuraciones no sería de tamaño excesivo. En tercero, que en la primera organización del procedimiento la salida del ordenador era demasiado grande para poder manejarla; en casos parecidos se repetía el mismo razonamiento con excesiva frecuencia. Cuarto, era evidente que el procedimiento contenía algunos fallos, pues había vértices de carga terminal positiva para los que no podía asegurarse la existencia de configuraciones geográficamente buenas. Finalmente, el programa generó una tremenda cantidad de información en tan sólo unas cuantas horas de funcionamiento, con lo que supimos que sería posible experimentar con frecuencia.

Tanto el programa como ciertos aspectos del proceso de descarga tuvieron que ser modificados para superar los problemas que las primeras pasadas de ordenador pusieron de manifiesto. Como nos fue posible mantener la estructura básica del programa, no resultó demasiado difícil efectuar estas modificaciones, y un mes más tarde realizamos un segundo conjunto de pasadas. Una vez corregidas las dificultades principales pudimos concentrarnos en otras más sutiles. Tras cierto tiempo de estudio también encontramos solución para estos problemas y volvimos a modificar el programa en consecuencia.

El diálogo entre hombres y máquinas se prolongó durante otros seis meses, hasta que estuvimos convencidos de que nuestro procedimiento permitiría obtener un conjunto inevitable de configuraciones geográficamente buenas en un intervalo de tiempo razonable. Fue en este momento cuando nos decidimos a demostrar formalmente que el procedimiento proporcionaría un conjunto inevitable de configuraciones geográficamente buenas. Teníamos que dejar de lado el trabajo experimental y describir el procedimiento total. Era necesario demostrar que se habían estudiado todos los casos y que aquellos casos no examinados por el programa de ordenador eran todo lo sencillos que parecían serlo.

Con gran sorpresa por parte nuestra, esta tarea resultó ser extraordinariamente difícil, invirtiéndose en ella más de un año. Fue necesario formular definiciones suficientemente generales de los términos, y demostrar proposiciones abstractas referentes a ellos. Tuvieron que examinarse con detalle todos los casos particulares, incluso los que en la práctica serán más inverosímiles, lo que frecuentemente requirió prolijo análisis. Finalmente, hacia el otoño de 1974, teníamos terminada una laboriosa demostración de existencia de un conjunto inevitable finito de configuraciones geográficamente buenas, y disponíamos de un procedimiento para construir tal conjunto, así como cotas precisas (aunque mucho mayores de lo deseable) del tamaño de las configuraciones del conjunto. El procedimiento que habíamos diseñado era para nosotros de la mayor importancia, pues nos proponíamos aplicarlo en la demostración del teorema de los cuatro colores. (Poco más tarde, Walter Stromquist, uno de los principales comunicantes de la teoría de reducción, y por entonces estudiante posgraduado en Harvard, construyó una elegante demostración de existencia de conjuntos inevitables de configuraciones geográficamente buenas. Sin embargo, como la demostración de Stromquist no proporciona ningún método de construcción efectiva de las configuraciones del conjunto, parecía poco probable que su demostración fuese inmediatamente aplicable a la conjetura de los cuatro colores.)

Cuando demostramos que nuestro procedimiento funcionaría para configuraciones geográficamente buenas desconocíamos todavía el grado de complejidad de la realización efectiva del proceso. Decidimos ensayarlo en el caso particular de triangulaciones carentes de pares contiguos de vértices de grado cinco. Se trata, evidentemente, de una fuerte restricción, pero el conjunto inevitable de configuraciones geográficamente buenas que obtuvimos fue francamente pequeño: tan sólo 47 configuraciones, y ninguna con anillo de tamaño mayor que 16. Sospechamos por ello que la solución del problema sin restricciones podría ser unas cincuenta veces mayor (lo que resultó ser una estimación algo optimista), y decidimos que había

buenos motivos para continuar por el mismo camino. A comienzos de 1975 volvimos a modificar el programa, para que esta vez produjese configuraciones libres de obstáculos en lugar de configuraciones geográficamente buenas, obligándolo a buscar conjuntos en los que la mayoría de configuraciones tuviesen pequeño tamaño anular. Al pasar el programa quedaron de manifiesto ciertos fallos, pero hubo también una sorpresa muy agradable: al sustituir la condición de que las configuraciones fuesen geográficamente buenas por la de que careciesen de obstáculos, el tamaño del conjunto inevitable tan sólo se duplicó.

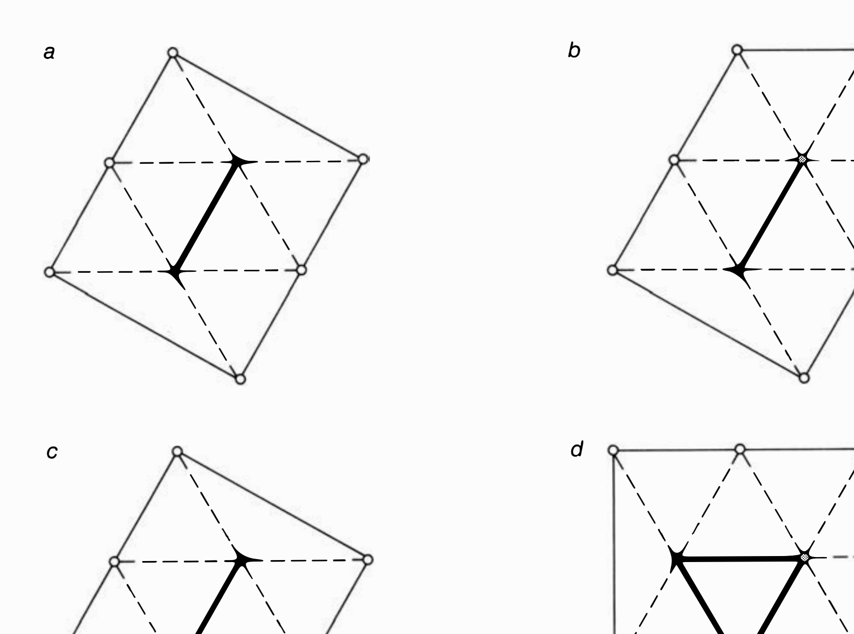
El programa, que hasta ese momento había absorbido todas nuestras ideas y perfeccionamientos, comenzó a sorprendernos. Al comprobar manualmente los análisis realizados por las primeras versiones del programa siempre habíamos podido predecir el camino que el programa iba a seguir; pero ahora el ordenador estaba actuando como una máquina de jugar al ajedrez, elaborando estrategias múltiples basadas en todos los recursos que le habían sido enseñados, y con frecuencia era mucho más diestro en su análisis de lo que nosotros hubiésemos sido. En cierto sentido, el programa exhibía su superioridad no sólo en la ejecución de las partes más mecánicas de la tarea, sino también en ciertas áreas de trabajo intelectual.

En el verano de 1975, estábamos convencidos de que había buenas posibilidades de poder hallar un conjunto inevitable de configuraciones, cada una de las cuales carecería de obstáculos y sería, verosímilmente, reducible. Tal conjunto contendría seguramente algunas configuraciones irreducibles, pero pensábamos que mediante alguna pequeña modificación de procedimientos podríamos sustituirlas por configuraciones reducibles. Tendríamos ahora que verificar, por primera vez desde el comienzo de nuestro trabajo, la reductibilidad de las configuraciones.

Comenzamos preparando un programa que efectuase la comprobación de reductibilidad en su interpretación más mecánica, trabajando con el lenguaje compilador del ensamblador IBM 360 de la Universidad de Illinois. Durante el otoño anterior se había unido a nosotros John Koch, estudiante postgraduado de informática que deseaba preparar una disertación sobre la reductibilidad de configuraciones de tamaño anular pequeño. (Frank Allaire, de la Universidad de Manitoba, y Edward Swart, de la Universidad de Rhodesia, estaban realizando trabajos similares sin nosotros saberlo.) En otoño de 1975, Koch había preparado ya programas capaces de comprobar la reductibilidad mecánica de configuraciones de tamaño anular máximo igual a 11, y progresaba hacia investigaciones más generales. Durante los meses siguientes modificamos con ayuda de Koch su trabajo sobre configuraciones de tamaño anular 11 con el fin de preparar programas para configuraciones de anillos con 12, 13 o 14 países. Finalmente, modificamos todavía más estos programas con el fin de introducir un proceso de reducción más general, diseñado por Birkhoff.

En este momento nuestro trabajo sobre el proceso de descarga llegó a punto muerto. Para mejorar el programa se requerían cambios estructurales y no meros ajustes de parámetros o pequeños añadidos, y cada cambio llevaría consigo una modificación fundamental del programa. Decidimos consiguientemente desechar el programa, y efectuar manualmente el proceso de descarga; se tendría de esta forma una flexibilidad mayor, y se podría, siempre que fuese necesario, modificar localmente el procedimiento. En diciembre de 1975, hicimos un descubrimiento que nos animó a proseguir. Una de las reglas que definían nuestro proceso de descarga era demasiado rígida. Al relajarla, la eficiencia del proceso mejoró considerablemente.

Parecía que con el nuevo proceso de descarga podríamos construir un conjunto inevitable cuyas configuraciones re-



PROCESO DE DESCARGA que genera un conjunto inevitable de configuraciones reducibles redistribuyendo la carga positiva de una triangulación arbitraria (sin vértices de grado menor que cinco) de forma que solamente aparezca carga positiva en las configuraciones reducibles. Como todo mapa contiene carga total positiva, se puede formar un conjunto inevitable eligiendo configuraciones que aparezcan en todo tipo de entorno de una carga positiva. Si todas las configuraciones del conjunto son reducibles no puede existir ningún mapa minimal pentacromático, y el teorema de los cuatro colores ha quedado demostrado. Se obtiene un proceso sencillo de descarga transfiriendo 1/5 de carga positiva desde cada vértice cargado positivamente hasta cada uno de sus vecinos negativamente cargados. En este proceso, un vértice de grado cinco *(círculos negros)* acaba con carga positiva solamente cuando tenga *(a)* un vecino de grado cinco, o *(b)* un vecino de grado seis *(círculos grises)*, con lo que se ve obligado a retener carga. Un vértice de grado seis nunca termina con carga positiva, porque teniendo carga inicial cero, nunca recibe carga positiva. Un vértice de grado siete termina con carga positiva solamente cuando tenga al menos seis vecinos de grado cinco; en tal caso, al menos dos de ellos serán adyacentes, como en *(a)*. Los vértices de grado mayor que siete nunca llegan a recibir carga suficiente para compensar su carga negativa inicial. El conjunto inevitable generado por este procedimiento de descarga está formado por un vértice de grado cinco unido por un lado a otro vértice de grado cinco *(a)*, y por un vértice de grado cinco unido por un lado a un vértice de grado seis *(b)*. Estas dos configuraciones no son reducibles. Si el procedimiento se modifica, transfiriendo 1/3 de carga unidad positiva desde cada vértice cargado positivamente hasta cada uno de sus vecinos con carga negativa, se obtiene un conjunto ligeramente mejor *(c, d)*. Si se transfiere 1/2 de unidad de carga, el conjunto resultante es parecido a uno de los que fueron obtenidos por los autores al comienzo de su trabajo en este tipo de procesos.

ducibles fueran más pequeñas que las producidas por procedimientos anteriores. El tiempo de ordenador requerido para efectuar la demostración sería probablemente menor de los que indicaban estimaciones precedentes. A pesar de todo, era evidente que aún para obtener un conjunto inevitable óptimo se necesitaría un considerable esfuerzo de cálculo. Edward F. Moore, de la Universidad de Wisconsin, había desarrollado un potente método para construir mapas carentes de configuraciones reducibles pequeñas. Por ejemplo, había creado un mapa donde el tamaño anular de su configuración reducible mínima era 12 *(véase la ilustración de la página 79)*. Así pues, todo conjunto inevitable de configuraciones reducibles ha de contener al menos una configuración de tamaño anular igual a 12. El trabajo de Moore proporciona una cota inferior del tamaño anular, aunque nosotros estamos ahora convencidos por completo de que es necesario que exista una configuración de anillo con 13 países, y que es muy probable que también sea necesario que exista una con 14. (Nuestra demostración prueba que ya no son necesarias configuraciones mayores.)

Comenzamos, en enero de 1976, la construcción de un conjunto inevitable de configuraciones reducibles utilizando nuestro nuevo proceso de descarga. La versión final de este proceso tenía una ventaja más con la que asegurar la reductibilidad de las configuraciones finales. El procedimiento era esencialmente auto-modificante; el programa estaba preparado para identificar áreas críticas, zonas problemáticas –configuraciones que aparentemente resistirían todo esfuerzo de reducción– y modificarse a sí mismo a fin de mover carga positiva a distinto lugar. Como el proceso de descarga se ejecutaba sin ayuda de ordenador, sabíamos que el proceso podría modificarse de acuerdo con las áreas críticas que encontrásemos.

Comenzamos haciendo una aproximación, una pasada preliminar de nuestro proceso de descarga. Examinamos todos los casos posibles en los que un vértice era forzado a ser positivo, y en cada caso se estudió su entorno para hallar una configuración libre de obstáculos. En el caso de no hallarse ninguna, el entorno recibe el nombre de entorno crítico, lo que implicaba que el proceso de descarga habría de ser modificado para evitar el problema. Sin embargo, aun cuando se hallase una configuración sin obstáculos, no se podía todavía garantizar que fuese una configuración reducible. Se usaron

MATASELLOS del departamento de matemáticas de la Universidad de Illinois, en Urbana-Champaign, en homenaje a la resolución del problema de los cuatro colores por dos de sus miembros.

los nuevos programas de reducción para intentar hallar en el entorno alguna configuración sin obstáculos que además fuese reducible. De no hallarse ninguna también se consideraba crítico a ese entorno.

Este método para desarrollar un conjunto inevitable de configuraciones reducibles (conforme se perfeccionaba el proceso de descarga) sólo fue posible tras un nuevo diálogo con el ordenador. Para determinar cuáles eran los entornos críticos fue preciso poder efectuar rápidamente la comprobación de reductibilidad, tanto en tiempo de ordenador como en tiempo real. Por fortuna, raramente fue necesario tener que esperar resultados más de unos cuantos días, aunque frecuentemente se precisó una considerable cantidad de tiempo de ordenador. Como esta intensa interacción entre el hombre y la máquina era indispensable en nuestro trabajo, debemos explicar las circunstancias que la hicieron posible.

Aunque en su momento las concesiones que se nos hicieron para utilización del ordenador nos parecían completamente naturales, hemos descubierto más tarde que tuvimos verdadera suerte al hallarnos trabajando en la Universidad de Illinois. En ella, la combinación de un gran instituto de cálculo y de una nada oscurantista política sobre el uso de ordenadores en investigación, nos proporcionó oportunidades que al parecer son casi inalcanzables en prácticamente ninguna otra universidad o centro de investigación. A pesar de no poder garantizar que nuestra labor condujese a demostrar el teorema de los cuatro colores, se nos dieron bastante más de 1000 horas de ordenador, en lo que, estamos convencidos, fue admirable muestra de fe en el valor de la investigación matemática pura. El centro de cálculo nos informó de que en vista de que los ordenadores de la universidad no estaban siendo utilizados por completo en todo momento por el trabajo de clase y otros tipos de trabajos de investigación, podríamos ser incluidos en un pequeño grupo de usuarios del ordenador a los que se permitía compartir los excedentes de tiempo muerto. Estamos seguros ahora de que tal concesión fue esencial en nuestro éxito.

Terminamos la construcción de un conjunto inevitable de configuraciones irreducibles en junio de 1976. El teorema de los cuatro colores estaba demostrado. Habíamos utilizado 1200 horas de funcionamiento de tres ordenadores distintos. El procedimiento de descarga definitivo se diferenciaba en unas 500 modificaciones de nuestra primera aproximación, modificaciones impuestas por el descubrimiento de entornos críticos. Para desarrollar el procedimiento fue preciso examinar manualmente unos 10.000 entornos de vértices de carga positiva y analizar mediante máquina más de 2000 configuraciones. Considerable parte de este material, incluida la reducción de 1482 configuraciones, se utilizó en la demostración final. Aunque el procedimiento de descarga (sin las reducciones) puede comprobarse manualmente en un par de meses, sería virtualmente imposible comprobar los cómputos de reducción en esta forma. De hecho, cuando sometimos nuestro artículo sobre la demostración al *Illinois Journal of Mathematics,* sus críticos comprobaron el procedimiento de descarga mediante nuestras notas completas, pero los cómputos de reductibilidad se comprobaron pasando un programa de ordenador independientemente del nuestro.

Muchos matemáticos, especialmente entre los formados antes de la aparición de los ordenadores de alta velocidad, se resisten a tratar al ordenador como instrumento matemático ordinario. Están convencidos de que un razonamiento es endeble cuando es imposible revisarlo, en todo o en parte, mediante cómputos manuales. Desde este punto de vista, la comprobación de resultados como el nuestro mediante programas de ordenador independientemente escritos no es tan convincente como la comprobación directa de las demostraciones. Las demostraciones tradicionales de los teoremas matemáticos son razonablemente breves y sumamente teóricas –cuanto más potente sea la teoría tanto más elegante es la demostración– y, por lo general, el mejor método para comprobarlas es revisarlas a mano. Pero incluso cuando la comprobación manual sea posible, en aquellos casos en que las demostraciones sean largas y contengan elevado número de cálculos, es difícil creer que la comprobación manual eliminará toda posibilidad de error. Además, cuando las comprobaciones son lo suficientemente rutinarias, como ocurre en nuestra demostración, probablemente sea más eficiente comprobar los programas de máquinas que comprobar cálculos manualmente realizados.

Si hay tantos matemáticos a los que desagradan las demostraciones largas, ello puede ser debido a que hasta recientemente tan sólo se han usado métodos que producen demostraciones breves. Ignoramos si podrá obtenerse o no una demostración breve del teorema de los cuatro colores. Se han anunciado varias nuevas demostraciones de longitud moderada, pero ninguna de ellas ha superado el escrutinio de los expertos. Aunque es concebible que alguna de estas demostraciones sea válida, también entra dentro de lo posible que las únicas demostraciones correctas sean las basadas en conjuntos inevitables de configuraciones reducibles, que, por consiguiente, exigirán cálculos de imposible verificación manual. Estamos convencidos de que existen teoremas de gran interés matemático que sólo podrán demostrarse mediante métodos computarizados. Aun cuando el teorema de los cuatro colores no fuera uno de ellos, proporciona un buen ejemplo de lo que podría hacerse para demostrar uno que sí lo fuera. No existe razón para pensar que no haya un gran número de problemas que exijan análisis diferentes, del tipo del nuestro.

Nuestra demostración del teorema de los cuatro colores sugiere que existe un límite para lo conseguible por métodos puramente teóricos incluso dentro de las matemáticas. De ella se deduce también que en el pasado se ha subestimado la necesidad de métodos computacionales en las demostraciones matemáticas. Es de gran valor práctico para los matemáticos, que pueden así determinar las capacidades y limitaciones de sus métodos. Tenemos la esperanza de que nuestro trabajo contribuya a avanzar en esta dirección, y que esta ampliación de las técnicas válidas de demostración justifique el inmenso esfuerzo dedicado durante los cien últimos años a la demostración del teorema de los cuatro colores.

Cómo empezó la Edad del Hierro

Hasta las postrimerías del segundo milenio a.C. el metal utilizado en el mundo mediterráneo fue el bronce. En pocos siglos se vio reemplazado por una nueva clase de metal: el acero

Robbert Maddin, James D. Muhly y Tamara S. Wheeler

Dice un viejo refrán que la necesidad es la madre del progreso. Aunque los abundantes inventos, no previstos, de los tiempos modernos parecen contradecirlo, la historia de la humanidad confirma en líneas generales la verdad del aforismo. Una investigación actual de historia de la tecnología nos proporciona un caso a propósito: la aparición, hace unos 3000 años, de una clase de acero que sustituyó al bronce. Este había sido hasta entonces el metal más utilizado por el mundo civilizado.

A lo largo de unos dos milenios, hasta el 1200 a.C. aproximadamente, las civilizaciones del Viejo Mundo satisficieron sus necesidades en relación al uso del metal –para herramientas, armas y otros útiles duraderos–, con varios tipos y aleaciones del cobre, bronce incluido. Hacia el fin de esta era, conocida comúnmente como la Edad del Bronce, las civilizaciones del Mediterráneo oriental sufrieron una serie de perturbaciones. No sabemos exactamente cuál fue la causa de las mismas, pero hay un factor evidente en los siglos siguientes: el uso del bronce disminuyó rápidamente y el uso del hierro –en concreto, el del hierro "acerado" o carbonatado– se incrementó cada vez con mayor rapidez.

Varias pruebas, necesariamente fragmentarias, pero lo suficientemente diversas como para inspirar cierto grado de seguridad, marcan el ritmo de la transición. El hierro era conocido como metal susceptible de ser trabajado, si no durante toda la Edad del Bronce, sí al menos durante gran parte de ella. Sin embargo, los yacimientos de la Edad del Bronce del Mediterráneo oriental y del suroeste asiático, representativos del momento histórico en torno al año 2000, han proporcionado un total de menos de 500 artefactos de hierro, de los cuales la gran mayoría son de uso ornamental. Los artefactos de bronce, procedentes de los mismos yacimientos, se cuentan por miles.

Este panorama contrasta con el número de cuchillos y armas hechas de ambos metales que se han hallado en los yacimientos griegos correspondientes al período comprendido entre el año 1050 y el 900 a.C. A. Snodgrass, de la Universidad de Cambridge, tabuló la abundancia relativa de bronce y de hierro de la manera siguiente: cuchillos de bronce ninguno; cuchillos de hierro más de 15; espadas de bronce una, espadas de hierro más de 20; puntas de lanza de bronce ocho, puntas de lanza de hierro más de 30.

Esta distribución tiene interés desde un doble punto de vista. En primer lugar, los artículos de bronce de esta clase pueden hacerse por fundición, método de producción rápido y fácil, mientras que los artículos de hierro similares había que manufacturarlos individualmente mediante la forja, un proceso arduo si lo comparamos con el anterior. En segundo lugar, el bronce, como su principal componente, el cobre, tiene una duración larga y puede rasparse y fundirse de nuevo, repetidamente. ¿Por qué los pobladores de estos yacimientos sustituyeron un proceso simple por otro más complejo? ¿Y qué sucedió con los artículos de bronce que habían fabricado en años anteriores?

Las futuras investigaciones arqueológicas deberán revelar por qué el bronce cayó en desuso. Entre las causas podía aducirse el colapso del comercio del estaño, el principal componente de las aleaciones en los artículos de la Edad del Bronce final. Por lo que respecta al rápido incremento del uso del hierro en sustitución del bronce, el panorama está bastante más claro. Tanto la investigación arqueológica como el estudio de las inscripciones antiguas han demostrado, en los últimos años, muchas cosas que eran aún conjeturas y han revelado muchas otras que eran enteramente desconocidas. Para presentar esta información en su contexto vamos a describir brevemente la metalurgia del cobre y del hierro tal como se conocía en los últimos milenios anteriores a la era cristiana.

El cobre se encontraba como un metal (en fragmentos de cobre nativo) y como un mineral que debía calentarse para producir metal. Para el propósito que nos guía podemos prescindir del cobre nativo. Los yacimientos más antiguos en los que tenemos documentada la producción de cobre están en Irán e Israel, y datan del quinto y cuarto milenio a.C. El mineral de cobre se encontraba en modestas cantidades en Chipre, en varias partes de Turquía, en Irán y en Israel. El mineral más común es la calcopirita, un complejo sulfuro de cobre y hierro que puede tostarse al aire libre para liberar el sulfuro y, después de ello, puede ya fundirse. Algunos minerales de cobre contienen arsénico: cuando se funden producen una aleación de cobre arsenical que puede calificarse de "bronce natural".

Los primeros que trabajaron el metal en el Mediterráneo oriental solían fundir el cobre llenando un horno de piedra con capas alternantes de carbón y mineral mezclado con fundente. El fundente servía para liberar, del mineral de cobre, los componentes que no se querían en el producto final. En conjunto, estos componentes se conocen con el nombre de ganga (palabra derivada originalmente del griego y aplicada a un filón mineral). Con el horno caliente, el fundente tiende a combinarse con la ganga y a separarla del metal. En muchos minerales del Mediterráneo oriental la ganga era sílice: cualquiera de los muchos óxidos silícicos. El fundente apropiado para estos minerales era el óxido de hierro hematites; el calor del horno combinaba el óxido de silicio y el

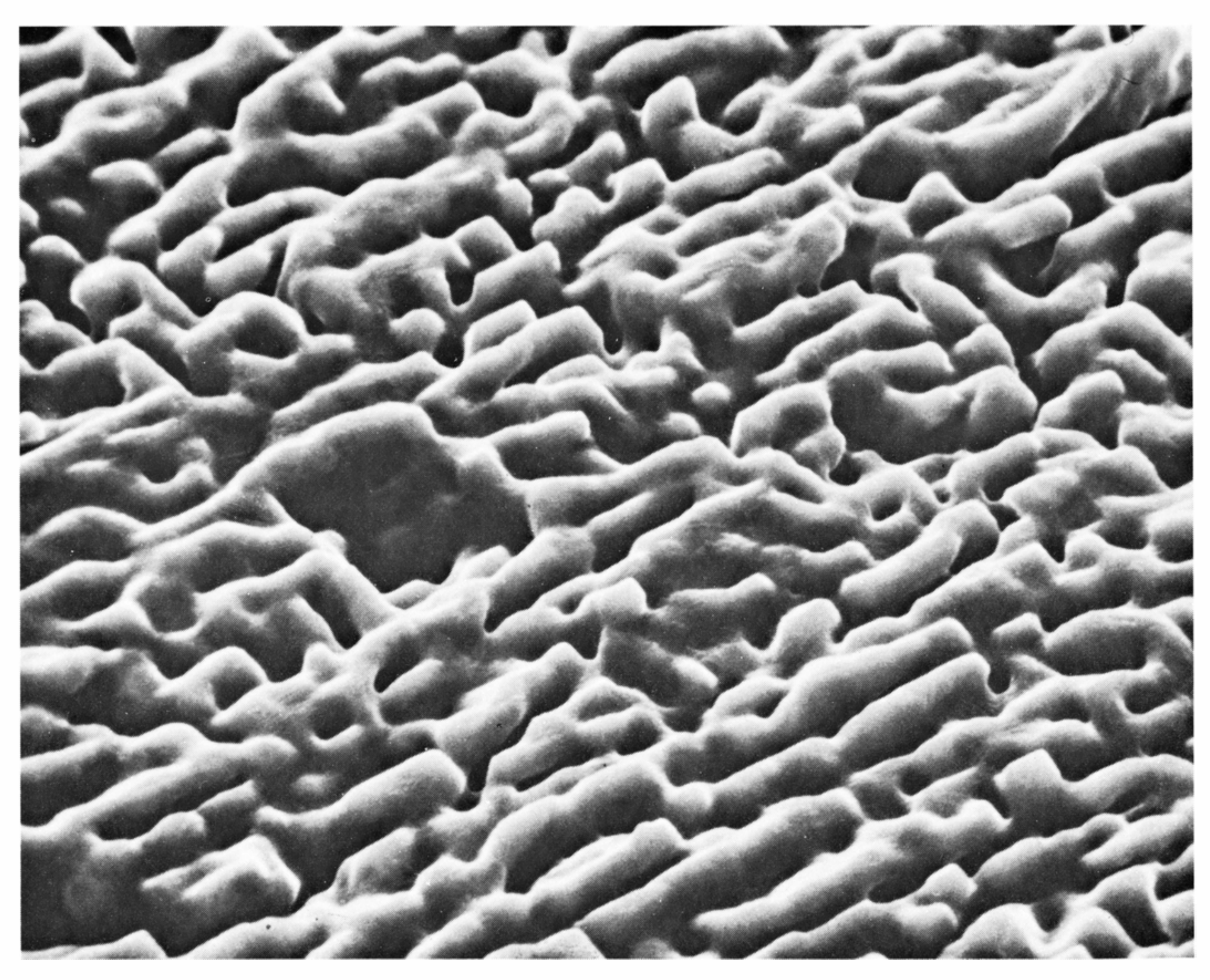
óxido de hierro formando un silicato férrico.

Cuando el horno estaba lleno, se prendía fuego al carbón vegetal. En algunos hornos el calor del fuego se incrementaba con una corriente de aire natural que se producía por un conducto de ventilación; en otros, el aire se impulsaba hacia la parte alta, hacia los lados o la parte inferior del horno a través de conducciones de arcilla. Conforme la carga del horno se calentaba, el carbón se oxidaba transformándose en monóxido de carbono; el gas caliente fluía hacia arriba, a través de la mezcla de mineral y fundente, sometiendo a ambos a una reducción química. Esta reducción se producía a unos 1100 grados Celsius y el cobre fundido se deslizaba hacia abajo quedando pudelado en la parte inferior del horno y dejando la ganga como una escoria. Los primeros metalúrgicos no podían prever ni medir con exactitud la temperatura del horno; seguramente, si no aparecía el cobre pudelado dejaban enfriar el horno y empezaban de nuevo el proceso con una carga distinta o con más aireación.

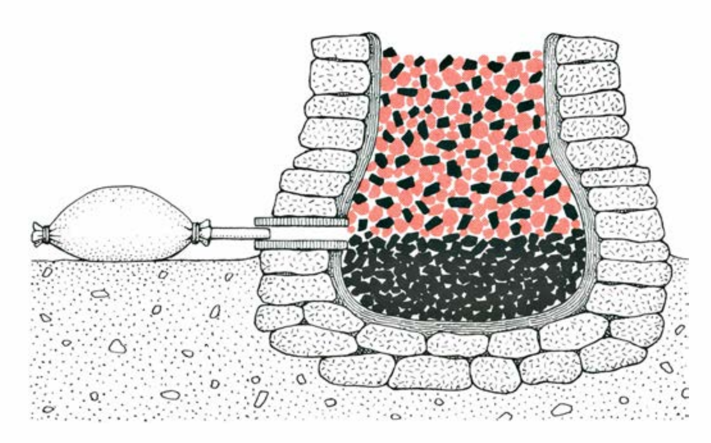
Si el mineral contenía un porcentaje más alto de arsénico, lo que hallaba el metalúrgico era un pudelaje de bronce natural. El bronce tenía la ventaja de ser más duro que el cobre. Sin embargo, si el pudelaje era meramente cobre puro blando, el metal fundido se podía endurecer martillándolo. Este modo de trabajar el cobre hizo posible que, durante el período anterior a la Edad del Bronce (cuya duración varía en las distintas partes del Mediterráneo oriental y del suroeste asiático), los que lo trabajaban pudieran manufacturar artículos de cobre de una duración considerable.

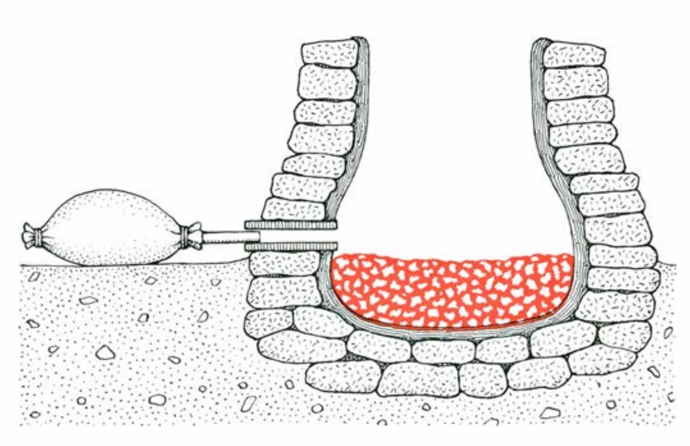
Con el tiempo, los minerales de cobre arsenical llegaron a fundirse extensamente, a causa de la mayor dureza del bronce natural. Más tarde, se descubrió que el cobre combinado con estaño, en lugar de arsénico, era también duro. Quizá fue la toxicidad del arsénico lo que impulsó a reemplazar bronce arsenical por bronce estañado. En el Mediterráneo oriental y el suroeste asiático el bronce estañado apareció por primera vez a principios del tercer milenio a. C.; en los primeros años del segundo milenio, la producción de bronce estañado superaba a la de bronce arsenical.

A finales del segundo milenio a.C. ocurrieron una serie de invasiones en el Mediterráneo oriental –algunas de ellas se atribuyen, en los textos egipcios de la época, a las hordas extranjeras conocidas colectivamente como "pueblos del mar"– que produjeron el colapso de la autoridad local en numerosas zonas. En los siglos que siguieron, descritos por los estudiosos del mundo antiguo como edad oscura, el hierro substituyó pronto al bronce como metal usado más comúnmente para herramientas,



ESTRUCTURA ESPONJOSA del hierro en lingotes, puesta de manifiesto en esta fotografía, sacada con microscopio electrónico de barrido. Presenta una muestra moderna de hierro fundido a una temperatura inferior a su punto de fusión. Para conseguir esta estructura esponjosa se redujo un óxido muy puro de hierro en una atmósfera de helio y monóxido de carbono; vemos el resultado aumentado en 2400 diámetros. Con una muestra mineral de hierro ordinario, los intersticios estarían rellenos de escoria, que podría extraerse mediante el martillado.



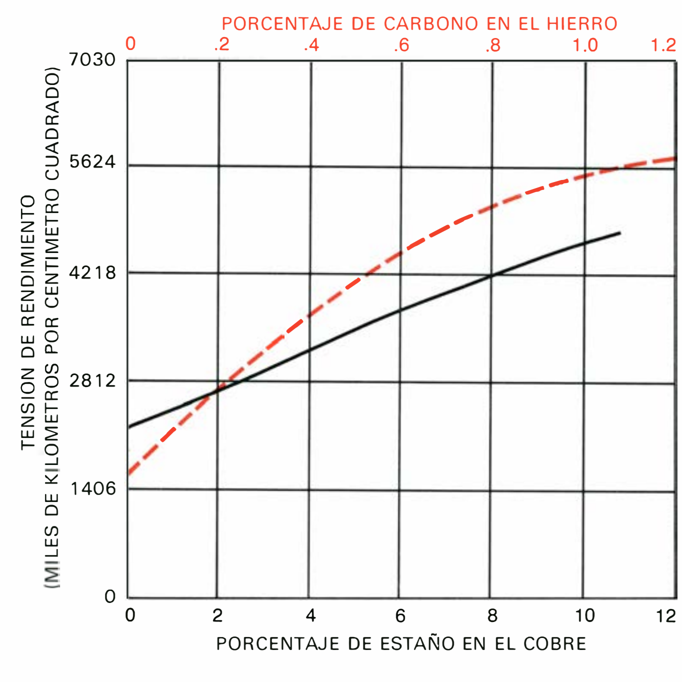


HORNO PARA FUNDIR HIERRO, según una reconstrucción ideal basada en los restos de los hornos de la Edad del Hierro en Europa. Primero se llenaba con una mezcla de mineral, que solía ser hematites o magnetita, y carbón vegetal *(izquierda)*. **Se encendía la carga, elevándose la temperatura del horno a unos 1200 grados Celsius mediante una corriente de aire impulsado por la tobera. Como el hierro no funde por debajo de 1537 grados Celsius, el producto del proceso de fundición** *(derecha)* **era una mezcla esponjosa de desperdicios no metálicos y hierro** *(color)* **conocida como zamarra. El herrero introducía de nuevo este lingote esponjoso en la fragua a unos 1170 grados Celsius transformando los desperdicios en una masa viscosa. El herrero separaba del hierro estos desperdicios mediante el martillado. Lo que quedaba era hierro dulce.**

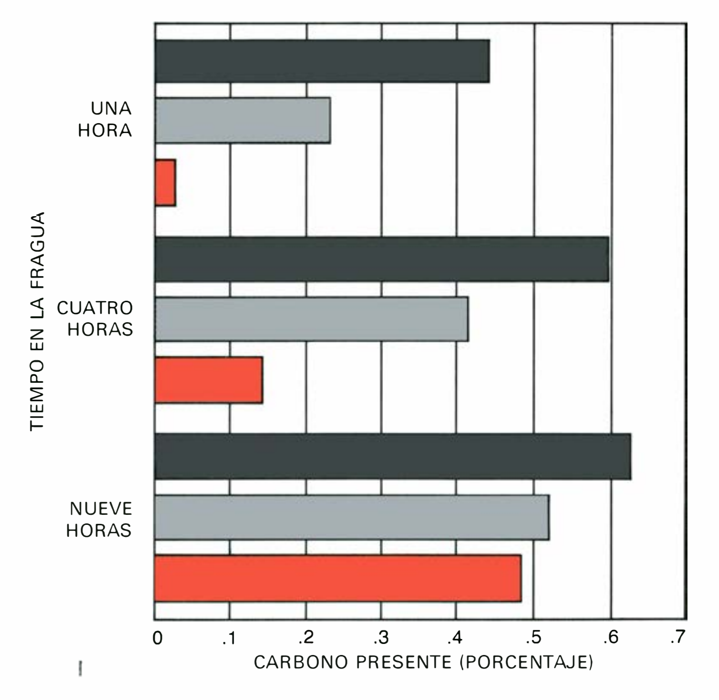
armas y otros instrumentos. Puesto que el bronce había sido satisfactorio para los mismos fines durante milenios y el hierro no se mostró utilitario, debe inferirse que el hierro no fue adoptado súbitamente a resultas de una innovación técnica sino a consecuencia de la repentina escasez del bronce. Esta, a su vez, se infiere de la interrupción del suministro de estaño, e incluso de cobre, a los fundidores de bronce del Mediterráneo oriental. De dónde procedía el estaño no lo sabemos. La fuente de suministro debía estar relativamente próxima a zonas mineras, como los Balcanes o el desierto oriental de Egipto; también pudieron ser regiones muy alejadas como Cornualles o el Irán oriental.

Los primeros que trabajaron el hierro lo producían a partir de minerales, sobre todo del hematites y la magnetita, mediante un proceso de fundición muy semejante al utilizado para producir cobre. Sin embargo, había una diferencia importante. El hierro no funde a temperaturas inferiores a los 1537 grados Celsius y la temperatura más alta que se podía alcanzar en la fundición primitiva era de unos 1200 grados. Si se funde el mineral de hierro a esta temperatura no se obtiene un pudelaje del metal sino una masa esponjosa mezclada con óxido de hierro y silicato férrico. Estas sustancias no metálicas, que unidas forman la escoria, proceden de la combinación del óxido ferroso y la ganga de sílice durante el proceso reductor.

La más común de las sustancias no metálicas es la fayalita, que permanece



EL EFECTO del proceso de conversión del hierro en acero es fortalecer el metal hasta que su dureza sea bastante mayor que la del bronce. La adición de un 1 por ciento de estaño al cobre produce una aleación de bronce *(negro)* **con una resistencia de 2192 kilogramos por centímetro cuadrado y adicionando un 5 por ciento de estaño resulta bronce con una resistencia de 3515 kilogramos por centímetro cuadrado. Si el hierro se le adiciona un 0,4 de carbono resulta una aleación carburada** *(color)* **con una resistencia de más de 3515 kilogramos por centímetro cuadrado. Un hierro que contenga un 1 por ciento de carbono es más de 7030 kilogramos por cm² más fuerte que un bronce que contenga 8 por ciento de estaño.**



LA CARBURACION DEL HIERRO resulta de la difusión del carbono en el hierro en un fuego de carbón vegetal. La proporción de la difusión depende del calor del fuego y del tiempo que el hierro está en la fragua. A 920 grados Celsius, la cantidad de carbono que se difunde a una profundidad de medio milímetro de la superficie del hierro en una hora *(barras superiores)* **es de un 0,5 por ciento por unidad de peso** *(negro)*, **pero sólo un 0,02 por ciento penetra a 1,5 mm de profundidad** *(calor)*. **Una exposición prolongada incrementa el porcentaje de carbono difundido a un mm de profundidad** *(gris)* **y a 1,5 mm de profundidad bajo la superficie del hierro, sin que se incremente mucho a 0,5 mm.**

viscosa a temperaturas inferiores a los 1177 grados Celsius. Por eso, los metalúrgicos sacaban del horno una masa esponjosa de hierro, la recalentaban en la forja y extraían cuidadosamente la fayalita mediante el martillado. Al mismo tiempo, el martillado transformaba el lingote de hierro poroso en una red continua de granos de hierro mezclados con algunas hebras de escoria que no había podido ser eliminada. La esponja o lingote era la materia prima del herrero: los artículos de hierro se hacían calentando la esponja y conformándola luego a golpe de martillo.

El herrero debía trabajar con un pobre substituto del bronce. El hierro esponjoso es un metal blando: su tenacidad o resistencia es de unos 2812 kilogramos por centímetro cuadrado, escasamente superior a la del cobre puro (2250 kilogramos por centímetro cuadrado). El proceso de endurecimiento que presenta la forja continua puede elevar la resistencia del hierro a casi 7030 kilogramos por centímetro cuadrado. Sin embargo, el bronce que contiene un 11 por ciento de estaño tiene una resistencia de 4218 kilogramos por centímetro cuadrado después de su fundición y de unos 8436 kilogramos por centímetro cuadrado después del trabajo en frío. El bronce es mucho mejor que el hierro esponjoso como material para la fabricación de armas y herramientas.

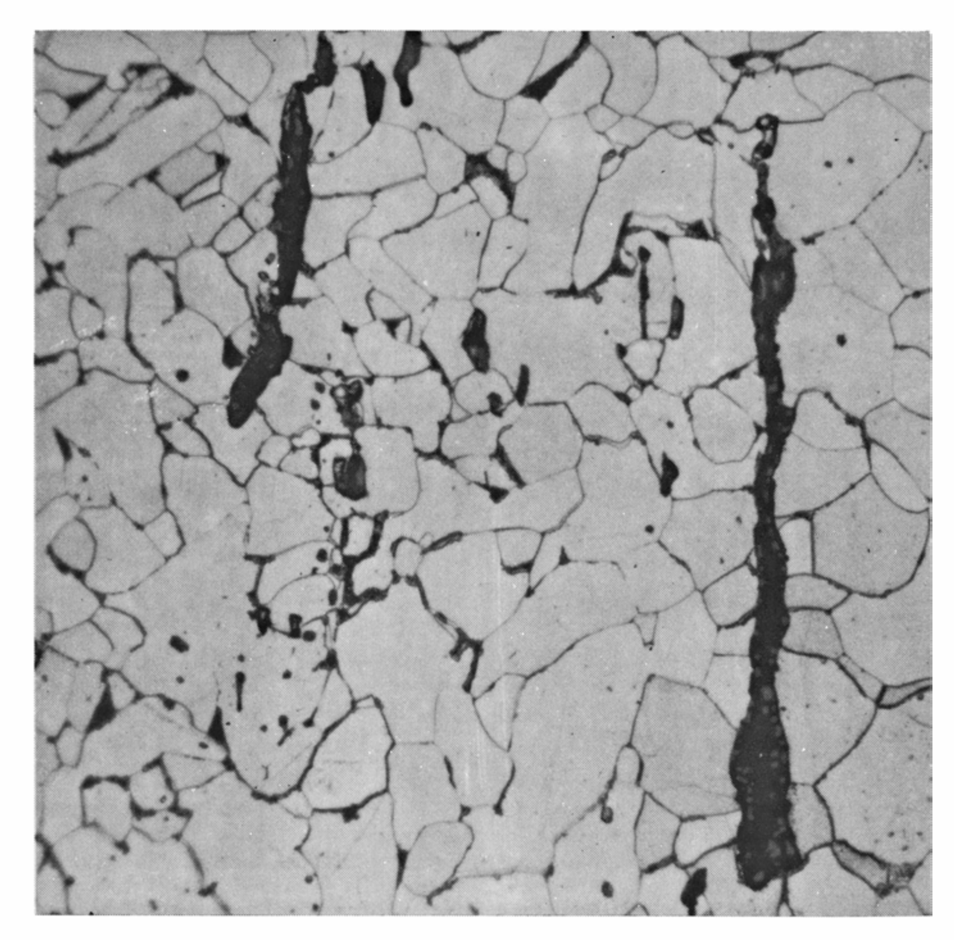
El bronce tenía además otras ventajas sobre el hierro. Poniéndolo al horno a las temperaturas que los antiguos metalúrgicos podían alcanzar, podía fundirse. Un bronce que contenga un 11 por ciento de estaño empieza a perder fluidez cuando se enfría a 1000 grados Celsius y está completamente sólido a 831 grados. Puesto que el hierro puro no funde por debajo de 1537 grados, no podía utilizarse para la fundición. Cuando se mezcla con grandes cantidades de carbón –digamos un 4 por ciento– funde a unos 1150 grados. Sin embargo, cuando se vuelve a solidificar se torna muy quebradizo. En cualquier caso, el hierro no se logró fundir antes de mediados del primer milenio a.C. cuando el proceso había sido ya ensayado por los chinos en el Lejano Oriente.

Como se deduce de los moldes que han llegado hasta nosotros, o como se infiere de las referencias escritas, los métodos de fundición usados en la Edad del Bronce eran notablemente variados.

Había moldes de una, de dos o de muchas piezas, hechos de tierra, arcilla y piedra. Algunos tenían núcleos; y el proceso de cera perdida permitía moldear el bronce en formas complejas y con superficies profusamente decoradas. El hierro,



LA PERLITA, una microestructura característica de los aceros al carbono, tiene, a gran aumento, una disposición estratificada. Esta fotografía sacada con microscopio electrónico de barrido aumenta la muestra en 1250 diámetros. Presenta una réplica, revestida de oro, de la superficie corroída de una espada de hierro acerado descubierta en un yacimiento de Israel: Tel Fara South. Aunque el hierro está oxidado, se conservan restos de microestructura de perlita. Dicha espada fue ofrecida, para su estudio, por el Instituto de Arqueología de la Universidad de Londres.



PUNZON DE HIERRO procedente de un yacimiento de Grecia, visto en sección longitudinal y aumentado 200 veces. Las zonas estrechas y oscuras son hebras: fragmentos de escoria del lingote de hierro primitivo que se deformaron, pero que no se extrajeron en el martillado que precedió a la forja del punzón. El Museo de Numismática de Atenas lo puso a disposición para su estudio.

excepto en el quebradizo estado en que posee un alto contenido de carbono, era menos útil para estos procesos, que producían artículos necesarios y gran número de objetos puramente ornamentales, como figurillas o joyas. El bronce poseía una última ventaja: se corroe lentamente y su pátina verde característica se consideraba decorativa. El hierro se corroe rápidamente, y en ese proceso suele sufrir daños considerables.

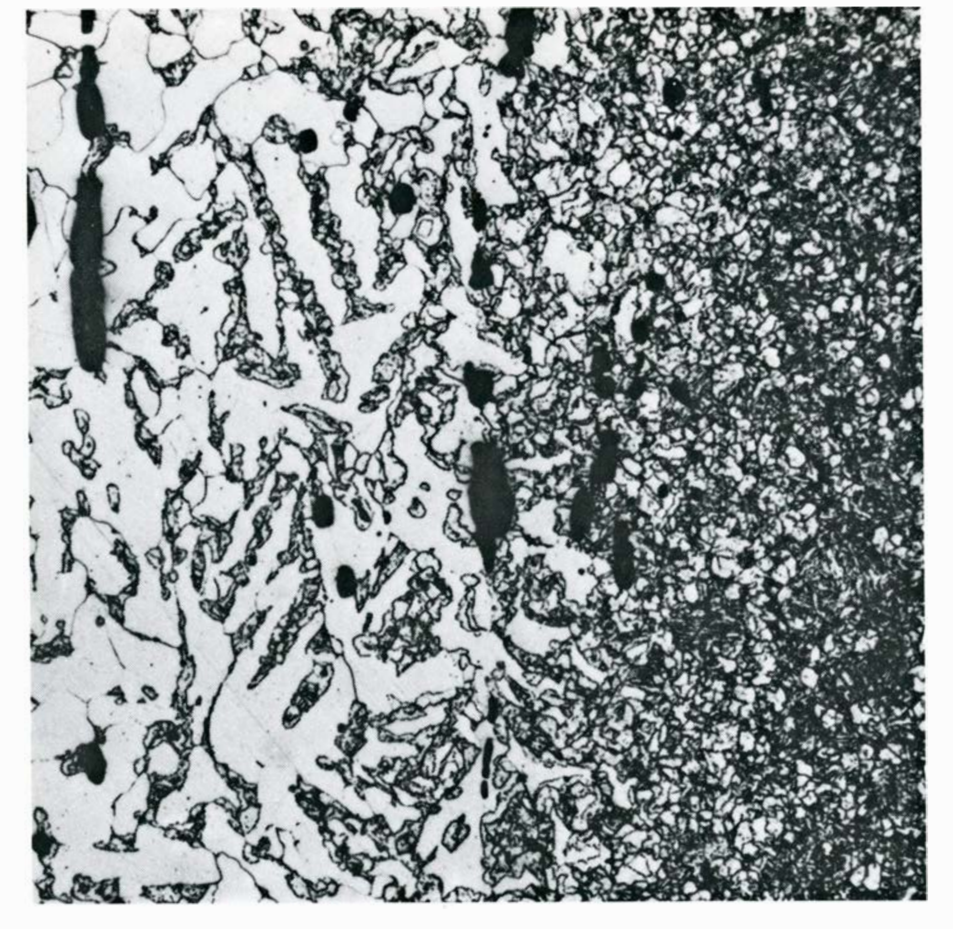
¿Cómo pudo, pues, convertirse el hierro en un substituto satisfactorio del bronce en el segundo milenio a.C.? La respuesta es que si el hierro esponjoso se trata de una determinada manera puede transformarse en una aleación, que, para muchos fines, resulta muy superior al bronce. Este tratamiento es el acerado (es decir, el proceso de conversión del hierro en acero), y su descubrimiento fue, a buen seguro, puramente accidental. Sucedió lo siguiente. Cuando el herrero recalentaba el lingote para separar la escoria, quemaba carbón vegetal en la fragua. Necesitaba calentar el lingote a unos 1200 grados Celsius para que la escoria fuera viscosa; probablemente, no dejaba enfriarse el fuego a menos de 800 grados hasta que terminaba el trabajo. El lingote estaba en contacto directo con el carbón al rojo blanco y con el monóxido de carbono caliente producido por la combustión. A esta temperatura, una pequeña cantidad de carbono procedente de ambas fuentes se difunde lentamente por el hierro, convirtiéndolo en acero al carbono a cierta profundidad de la superficie.

El tiempo requerido para la difusión del carbono por el hierro sigue unas leyes físicas simples. Por ejemplo, si establecemos una gráfica de concentración de carbono y profundidad de penetración a 950 grados Celsius de temperatura, vemos que, a las nueve horas, la concentración a 1,5 mm de profundidad es de un 0,5 por ciento. A temperaturas mayores, los átomos de carbono se difunden en el hierro más rápidamente: a 1150 grados Celsius, al cabo de nueve horas, la concentración a la misma profundidad puede llegar a un 2 por ciento.

En términos de metalurgia moderna, el hierro carbonatado a una temperatura de más de 910 grados Celsius tiene la microestructura de la forma del acero conocida como austenita. Por debajo de los 727 grados Celsius, la austenita se fracciona en dos componentes: la ferrita o hierro puro y el carburo de hierro, conocido como cementita. A tal reacción se la denomina eutectoide. Esta descomposición bifásica da como resultado la microestructura de la forma de acero conocida como perlita, que alterna capas de ferrita y cementita. La proporción de carbono en el metal se refleja en esta microestructura. Si el hierro no contiene carbono no habrá perlita. Si la aleación contiene un 0,8 por ciento de carbono, la microestructura que se forme contendrá un 100 por ciento de perlita.

Es ilustrativo comparar la capacidad de resistencia del hierro carbonatado y del bronce. Un contenido en carbono que oscile entre un 0,2 por ciento y un 0,3 por ciento da un hierro acerado de resistencia igual a la del bronce no trabajado: alrededor de 4218 kilogramos por centímetro cuadrado. Si el contenido en carbono es de un 1,2 por ciento el acero tiene una resistencia de 9842 kilogramos por centímetro cuadrado, que es algo mayor que la del bronce trabajado en frío. Si el herrero martillea luego en frío el hierro acerado, su resistencia se incrementa hasta 17.225 kilogramos por centímetro cuadrado, más del doble de la resistencia del bronce trabajado en frío.

El descubrimiento accidental del proceso que convierte el hierro en acero debió incitar a la experimentación porque, a su debido tiempo, los primitivos herreros pudieron controlar el proceso lo suficientemente bien como para desarrollar en el metal una serie de propiedades adecúadas a la función del objeto que estaban haciendo. Futuros estudios metalúrgicos (algunos están ya en proceso de elaboración) nos ayudarán a conocer el avance gracias al cual se pasó de la conversión accidental del hierro en acero a la intencionada. Por ejemplo, el carbono puede incorporarse al hierro por medios distintos de la carburación. Puede quedar retenido en los poros del lingote de hierro; si se forja, el carbono formará hebras en el metal. En una sección metalográfica pulida de una pieza convertida en acero accidentalmente, las hebras aparecerían como líneas irregulares. Una carburación totalmente accidental, por exposición al carbón vegetal en el curso del calentamiento y la forja, debe dejar alguna traza, aunque es más difícil detectarla microscópicamente. El contenido en carbono será probablemente bajo y la concentración irregular. Sin embargo, cuando encontra-



LA HABILIDAD METALURGICA en el siglo IV a.C. aparece en esta micrografía. Es una sección, vista a 300 diámetros de aumento, de una hoja de azuela hallada en Al Mina, las ruinas de una colonia comercial griega de la costa turca, cerca de Siria. El herrero que la hizo utilizó una lámina de hierro sometido a carburación para la parte de la azuela destinada al uso, pero economizó en la otra parte con hierro maleable. La micrografía, al mostrar la línea divisoria entre los finos granos de hierro acerado y los granos groseros de hierro sin carburar, habla de la maestría del herrero.

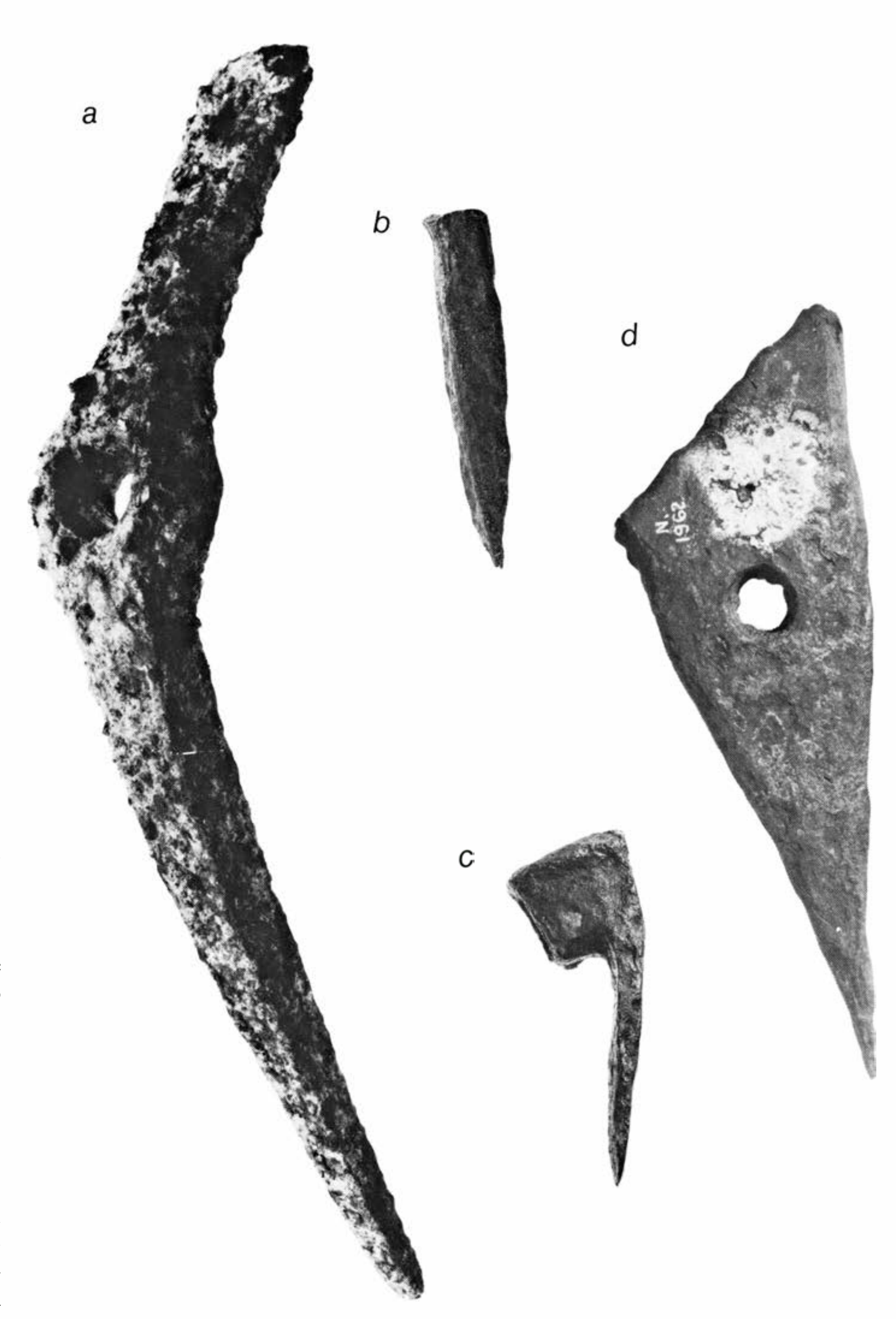
mos un objeto de hierro cuyas capas presentan porcentajes disimilares de carbono estamos ante una prueba incontrovertible de conversión en acero deliberada. El herrero no tendría motivo para hacer un objeto tal, a menos que conociera las diferentes propiedades de las distintas capas. El primer útil conocido de este tipo es un cuchillo egipcio, forjado probablemente entre el 900 y 800 a.C.

Sobre la base de nuestras investigaciones y de otras, parece evidente que en los inicios del siglo X. a.C. los herreros convertían conscientemente el hierro en acero. Muy pocos objetos de hierro del siglo XII a.C. o anteriores han sido analizados metalúrgicamente, por lo que el predominio de dicho proceso en ese período permanece bastante obscuro. No obstante, un cuchillo del siglo XII, procedente del yacimiento chipriota de Idalion, fue sometido a carburación, sin lugar a dudas, para aumentar su dureza. Un yacimiento de Monte Adir, en el norte de Israel, proporcionó un pico de hierro asociado a cerámica del siglo XII. Se pudo dudar de sacar una muestra para el análisis de la pieza, pero fue posible examinar la punta del mismo para calibrar su dureza. Las lecturas de las mediciones dieron un promedio de 38 en la escala de dureza de Rockwell "C", que es el característico del acero endurecido moderno.

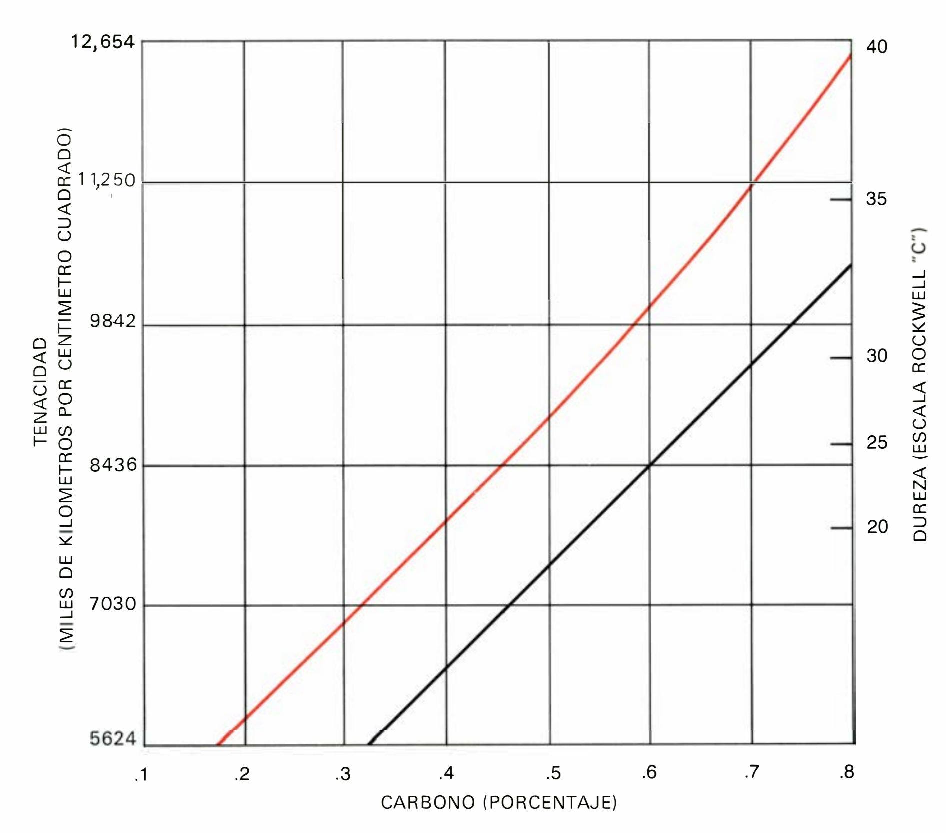
A diferencia del pico de Monte Adir, muchos de los artefactos de hierro antiguos que se han conservado en los yacimientos del Mediterráneo oriental están muy corroídos, hecho que complica el estudio metalúrgico. Sin embargo, hay varias técnicas que permiten detectar la perlita incluso en objetos totalmente oxidados. En una de estas técnicas, se prepara la superficie pulida para su estudio al microscopio, a más de 1000 aumentos. Puesto que las capas de ferrita de la perlita difieren de las capas de cementita en su composición química, a veces resulta posible detectarlas incluso en el óxido de hierro. El uso del microscopio electrónico de barrido mejora esta técnica. En este caso, no se examina la superficie pulida sino una fina película chapada en oro que es una réplica de la misma. Las capas de ferrita son más blandas que las de cementita, y ello incluso en artefactos sujetos a corrosión; aquéllas sobresalen claramente cuando la réplica se inclina para acentuar los distintos niveles de ambos componentes. La presencia de perlita en un artefacto de hierro prueba nítidamente que éste ha sufrido un proceso de carburación.

A partir del 900 a.C. la producción de utensilios de hierro se incrementó rápidamente. Aunque en esta época el estaño se halla de nuevo disponible en el Mediterráneo oriental, el bronce no reemplazó al hierro. Los yacimientos fechados entre los siglos X al VI a.C. han proporcionado grandes lotes de utensilios de hierro. En Hasanlu, al noroeste del Irán, una expedición de la Universidad de Pennsylvania sacó a la luz millares de armas de hierro. En Gordion, capital de la antigua Frigia, otra expedición de la Universidad de Pennsylvania descubrió una de las mayores colecciones de artefactos de hierro hallados jamás en el Mediterráneo oriental. En Nimrud, Irak, las excavaciones de Max Mallowan, de la Universidad de Oxford, rescataron un lote de objetos de hierro aún mayor.

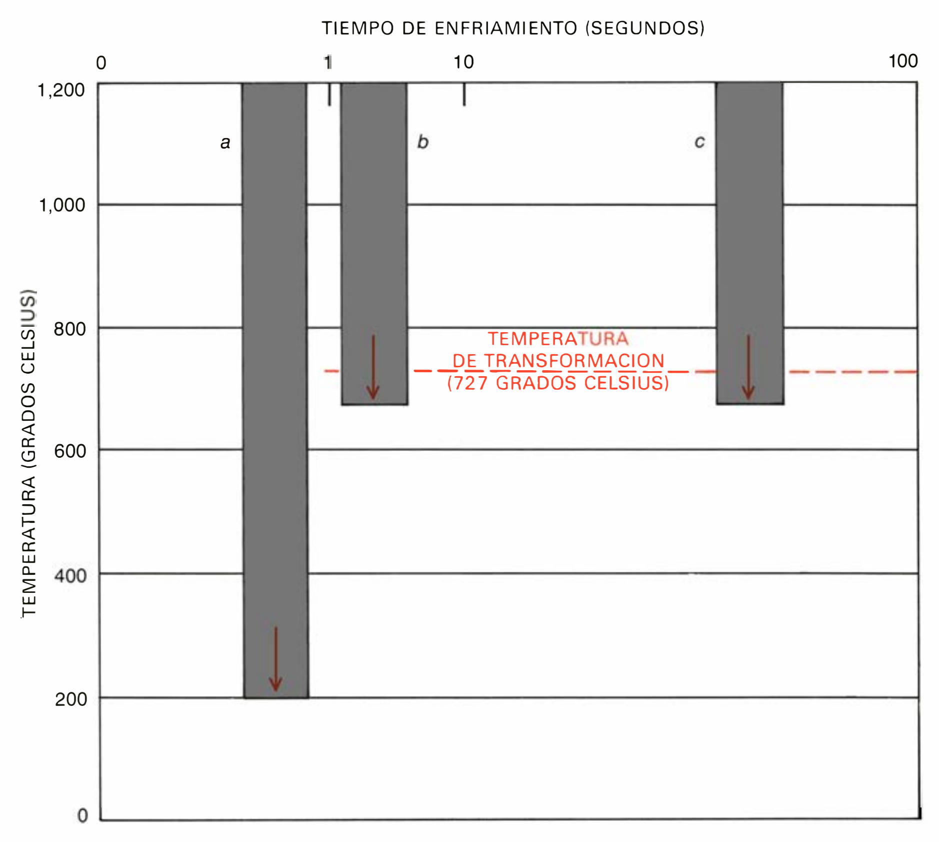
La prueba de la creciente populari-



CUATRO UTENSILIOS DE HIERRO hechos en el Próximo Oriente procedentes de yacimientos cuya datación oscila entre finales del segundo milenio a.C. y finales del primer milenio a.C. El mayor *(a)* es un pico de 39 cm, del norte de Israel, que se halló con cerámica del siglo XII a.C. Su punta, cuya dureza se sometió a prueba, dio una lectura media característica del acero endurecido moderno. El escoplo *(b)*, de Al Mina, mide casi 13 cm de largo; se le templó diestramente para conseguir que la matriz fuera dura y el filo cortante no resultara quebradizo. La pequeña azuela *(c)* de Al Mina, mide casi 12 cm de largo; probablemente, se la hizo a partir de la misma pieza de hierro que el escoplo y quizá por el mismo herrero. Podemos ver su microestructura en la ilustración precedente. La cuarta herramienta *(d)*, de 25 cm de largo, procede de las ruinas de Nimrud. Examinada al microscopio, muestra la estructura propia de la perlita en el hierro carburizado.



LA RESISTENCIA Y LA DUREZA del hierro sometido a carburación se incrementan en la medida en que aumenta el porcentaje de carbono en la aleación. Aquí, los grados de dureza se miden en la escala Rockwell "C". Si la velocidad de enfriamiento no es rápida, la microestructura de acero que se desarrolla es perlita grosera *(negro).* **Independientemente de su contenido en carbono, el acero de perlita grosera dura menos que el de perlita fina** *(color),* **como se desprende de la figura.**



EL EFECTO DEL TEMPLE en la microestructura del hierro sujeto a carburación aparece en este gráfico en función del período de tiempo que se requiere para rebajar la temperatura del hierro de su calor de forja (unos 1200 grados Celsius). Si el enfriamiento se produce en menos de un segundo *(a)*, **la microestructura que se forma es martensita, dura aunque un tanto quebradiza. Si se requieren hasta tres segundos, la microestructura resultante es perlita fina** *(b).* **Un enfriamiento prolongado, de más de diez segundos: formará una tercera clase de microestructura, la perlita grosera** *(c).*

dad del hierro va más allá de los hallazgos arqueológicos. Testimonios escritos neo-asirios y neo-babilonios hablan de un mundo totalmente distinto, desde un punto de vista tecnológico, de la civilización del período anterior al 900 a.C. Hacen referencia a hachas, azadones, picos, sierras, puntas de flecha, tijeras, grilletes e incluso mobiliario y lucernas, todo ello de hierro. Fue también el metal escogido para cuchillos y dagas. Un pasaje del corpus de la "literatura sapiencial" babilónica nos ofrece una incisiva nota a este respecto: "Una mujer es una trampa, un hoyo, un foso; una mujer es una aguda daga de hierro que degolla al hombre".

No hemos mencionado aún un segundo proceso que incrementa significativamente la calidad del hierro sometido a carburación. Se trata del temple: es decir, del rápido enfriamiento de una pieza caliente del metal sumergiéndola en agua. Un objeto de hierro acerado que se haya dejado enfriar por sí solo al aire libre desarrolla una microestructura de perlita muy grosera. En cambio, si el herrero agita el objeto terminado en el aire, acelerando el proceso de enfriamiento, la microestructura de perlita es más fina. Pero si el enfriamiento es más rápido, mediante el temple, puede impedir totalmente la aparición de perlita. El material acerado tiene una estructura completamente diferente, que se conoce con el nombre de martensita. La martensita es mucho más dura que la perlita, aunque tiene el inconveniente de ser quebradiza.

Si plasmamos gráficamente los resultados del proceso del temple, observamos que distintas tasas de enfriamiento –desde la temperatura del horno o de la forja, unos 1200 grados Celsius, hasta la temperatura de transformación, unos 700 grados– dan lugar a microestructuras diferentes. En un hierro que contenga un 0,8 por ciento de carbono, si el período de enfriamiento es de unos 60 segundos, se formará una microestructura grosera de perlita. Para que ésta sea fina, el período de enfriamiento deberá durar dos o tres segundos. Para producir martensita, hay que reducir la temperatura por debajo de los 220 grados Celsius en menos de un segundo; sólo el temple puede enfriar el material tan de prisa. Podemos suponer que, para objetos pequeños, verbigracia, puntas de flecha, la dureza era importante y podía tolerarse su carácter quebradizo. Si la mayoría de esos objetos pequeños se hacían de hierro con un 0,8 por ciento de carbono, un rápido templado podía convertirlos en martensita. Lo mismo podía suceder

con las capas superficiales de objetos de hierro mayores, en los cuales importaba la dureza y se toleraba la fragilidad, aun cuando el interior del objeto continuara siendo perlita, debido a que perdía calor más lentamente.

No se puede determinar cuándo se inventó el temple. Como ocurrió con el proceso de conversión del hierro en acero, es presumible que su descubrimiento fuera fortuito. Sin embargo, una prueba literaria indica claramente que los herreros del Mediterráneo oriental estaban familiarizados con este proceso en el siglo VII u VIII a.C. El pasaje corresponde al libro noveno de la *Odisea*. Atrapados en la cueva de Polifemo, el gigante de un ojo, Ulises y sus hombres consiguen emborracharlo y deciden cegarlo tomando del fuego un tronco de olivo ardiendo, hincándoselo en el ojo: "Así como el herrero, para dar el temple que es la fuerza del hierro, sumerge en agua fría una gran hacha o una azuela que produce un fuerte sonido sibilante, embrujándola, de igual manera chirriaba el ojo del Cíclope en torno de la estaca de olivo".

Esta descripción sólo pudo haber sido escrita por alguien que había visto un herrero templando hierro candente y que sabía que el temple incrementaba la dureza del metal. También sugiere que el endurecimiento al temple era algo nuevo en el mundo griego en ese momento. "Embrujándolo" es una traducción de la palabra griega *pharmasso;* implica que el hierro había sido tratado de alguna manera mágica como por medio de drogas o pócimas. Quizás Homero quedó perplejo ante el hecho de que el agua, que reblandecía e incluso disolvía muchos materiales, transformase el hierro sometido a carburación en un metal más duro que ningún otro conocido con anterioridad.

Es más difícil obtener pruebas del temple deliberado que del proceso consciente de conversión del hierro en acero. Por una parte, el temple pudo haberse descubierto al querer enfriar rápidamente un objeto forjado, quizá para utilizarlo inmediatamente o tal vez para evitar el tener demasiados objetos calientes en la herrería. Por otra, en los objetos de cierto tamaño la zona que corresponde a la martensita es la exterior, y esta capa es la que desaparece, por corrosión, con el paso de los siglos. A pesar de todo, existe alguna prueba, aunque pobre, de la existencia del temple deliberado. Por ejemplo, se ha analizado metalúrgicamente hasta 11 objetos procedentes de Nimrud. Aunque en todos ellos la corrosión ha sido muy tenaz, cinco de ellos muestran posibles indicios de temple. El resto presenta síntomas de carburación, pero no de temple.

Directamente del temple surge una tercera técnica de trabajar el hierro. Se trata del proceso de revenido, práctica que reduce la friabilidad inducida por el temple. Los herreros de la antigüedad debieron apercibirse pronto de que el temple hacía que los objetos fueran más quebradizos. El temple debía producir grietas en muchos objetos y los usuarios debían quejarse de las roturas. El revenido consiste en un recalentamiento, pero nunca por encima de la temperatura de transformación (727 grados Celsius) y afecta al carburo de hierro que el temple introduce en la microestructura del hierro, originando la martensita. El carburo precipita y luego entra en coalescencia por difusión. La temperatura exacta obtenida y el tiempo a que se somete el objeto a dicha temperatura determinan la cantidad de carburo de hierro que entra en coalescencia y, por consiguiente, la dureza y ductilidad final del metal. Conforme aumenta la ductilidad, decrece la dureza.

Probablemente, los herreros de la Antigüedad nunca procedieron intencionadamente al revenido. No hay cambio de color del hierro caliente en la gama crítica de temperaturas, y, por tanto, no había modo de graduar correctamente el calor en la fragua. A principios del siglo IV a.C. se desarrolló un método que subsanaba esa dificultad de conseguir un revenido auténtico y producía utensilios de hierro sujeto a carburación que eran fuertes y duraderos. Demuestra esta afirmación un escoplo de picapedrero hallado en la colonia comercial griega de Al Mina, yacimiento de la costa turca, cerca de los confines de Siria. El escoplo tiene una matriz compuesta en su mayor parte por martensita, si bien contiene nódulos de perlita cuya densidad se incrementa hacia la punta de la herramienta, y no hacia el interior como podía esperarse. Los ensayos de laboratorio indican que esta distribución inversa se consiguió recubriendo la punta de la herramienta con algún tipo de material aislante, arcilla por ejemplo, calentando la pieza y luego templándola. La arcilla debía romperse tan pronto como la pieza entraba en contacto con el agua, pero frenaba la velocidad del enfriamiento lo suficiente como para reducir la cantidad de martensita formada en el extremo cortante de la herramienta. Evidentemente, se seguían procedimientos tan sencillos y baratos como el empleado por el picapedrero de Al Mina para forjar su escoplo con un cuerpo sólido y un filo duradero.

Se halló en Al Mina otro ejemplo de innovación en el terreno de la forja de metales. El análisis de los elementos presentes en una azuela descubierta con el escoplo indica que ambas herramientas fueron hechas con el mismo lingote de hierro y quizá por el mismo herrero. La azuela se formó con dos láminas de hierro juntas martilladas en caliente. Una de ellas estaba sometida a carburación, la otra no. Las dos hojas combinadas recibieron forma por un proceso de forja y una vez terminada la azuela se la enfrió al aire y no al temple. Si el herrero era el mismo hombre que hizo el escoplo es evidente que se dio cuenta de que una azuela para trabajar la madera no necesita tener la dureza que precisa el filo de un escoplo para trabajar la piedra. Ajustó el proceso de manufactura, en consecuencia, con ahorro de tiempo y esfuerzo, probablemente. El momento en que por primera vez los herreros fueron capaces de controlar el proceso del trabajo del hierro de modo que las propiedades de los productos se adecuaran al fin a que estaban destinados, constituyó, sin duda, un hito en la historia de la tecnología. Ese instante puede muy bien ser el que nos indican los hallazgos de Al Mina.

En resumen, a comienzos del siglo VII a.C. como fecha más tardía, los herreros del Mediterráneo oriental dominaban dos de los procesos que convierten el hierro en un material útil para armas y herramientas: la carburación y el temple. Y a principios del siglo IV como máximo, hallaron un método para superar los inconvenientes del acero quebradizo, preservando, sin embargo, la ventaja de su dureza. Los intentos y ensayos de los herreros duraron cientos de años de experimentación y supusieron incontables horas en la fragua. Todos los artesanos de la Antigüedad –ceramistas, albañiles, picapedreros, tejedores, carpinteros y trabajadores del bronce, el cobre y los metales preciosos– tuvieron que enfrentarse con su trabajo de la misma forma empírica que el herrero, pero ninguno se encontraba ante una tarea tan formidable. Convertir el hierro en algo fuerte y duradero precisa complejos tratamientos. El herrero quizá no entendía, al menos al principio, por qué estos tratamientos mejoraban el hierro, pero su tenacidad, su conocimiento pragmático y su habilidad permitieron que, en los siglos que siguieron al fin del segundo milenio a.C., las gentes del Mediterráneo oriental pudieran dar el monumental salto que va de la Edad del Bronce a la Edad del Hierro.

Alucinaciones

Estas falsas percepciones, que pueden darse en cualquiera de nuestros sentidos, parecen ser iguales de una persona a otra. Todo indica que hunden sus raíces en las excitaciones del sistema nervioso central

Ronald K. Siegel

Un motorista que conduzca solo, por la noche, en un estado de extrema fatiga, puede fácilmente percibir cosas que no están en la realidad de su entorno: personas, animales, vehículos o formas extrañas. Tal percepción se denomina alucinación. Aunque la definición de la palabra (que viene del latín *hallucinari*, que quiere decir parlotear, soñar o delirar) dista mucho de tener una significación precisa, en psiquiatría se acepta como: "Una percepción sensorial falsa, en ausencia de un estímulo externo real. Puede inducirse por factores emocionales y por otros diversos, como son las drogas, el alcohol o la fatiga. Puede darse en cualquiera de los sentidos".

Según esta definición, cualquiera de nosotros las ha tenido alguna vez. Los exploradores solitarios, los aislados cazadores del Artico y los prisioneros en celdas de castigo han dado pruebas de haberlas experimentado. Algunas personas buscan profundizar en la experiencia mediante el consumo de drogas alucinógenas. En condiciones sociales adecuadas, estas percepciones pueden considerarse válidas. Juana de Arco fue santa debido a sus visiones, y los destellos luminosos que percibieron los astronautas se tomaron con toda seriedad. (En realidad estaban causados por rayos cósmicos.) Por otra parte, las instituciones correctivas aplican valoraciones negativas a percepciones parecidas de los presos. Sin embargo, todo este tipo de informes son necesariamente subjetivos; cuando alguien tiene una alucinación, la tiene solo, en la intimidad de su propia mente.

¿Tienen algo que ver las alucinaciones de una persona con las de otra? Con mis colegas del Instituto Neuropsiquiátrico de la Universidad de California, en Los Angeles, me propuse responder a esa cuestión por medio de experimentos. Hallamos que las alucinaciones tenían muchas cosas en común. Además, los experimentos apuntaban a los mecanismos subyacentes del sistema nervioso central como la fuente de origen de una fenomenología universal de las alucinaciones.

Una de las primeras clasificaciones de las alucinaciones fue la introducida, en 1853, por el francés Brierre de Boismont. Halló que las alucinaciones que se daban en estados de enfermedad, delirium tremens, intoxicación química, disfunciones nerviosas, pesadillas, sueños, éxtasis y fiebre, se caracterizaban todas por la excitación y producción de imágenes de la memoria y la imaginación. Su compatriota Jacques Moreau describió las alucinaciones como semejantes a los sueños en los que los estímulos visuales, auditivos y táctiles, imaginados, parecen ser reales. Anticipándose a lo que la investigación neuropsiquiátrica posterior descubriría, mantuvo que las alucinaciones provenían de la excitación del cerebro. La técnica de Moreau, que él describió en 1845, consistía en tomar hashish, lo que le permitía a la vez estar en estado de alucinación y ser capaz de explicar sus experiencias. (Moreau intentó persuadir a sus colegas médicos y amigos para que tomaran hashish; éstos se mostraron reacios, ya que consideraban que la ingestión de droga no era una forma aceptable de experimentación objetiva.

Los artistas bohemios y los escritores del París del siglo XIX eran más abiertos. Uno de ellos, el novelista Théophile Gautier, organizó el club de los adictos al hashish, entre cuyos miembros destacaban Balzac, Baudelaire, Dumas hijo y Victor Hugo. Algunos de los escritos de los miembros del club dan testimonio de la riqueza de las imágenes inducidas por el hashish.)

En Alemania y en los Estados Unidos, los primeros estudiosos de las alucinaciones siguieron un procedimiento similar de autoexperimentación y se centraron principalmente en las alucinaciones visuales. Empleando la mescalina, un alcaloide alucinógeno derivado del peyote, *Lophophora williamsii*, Heinrich Klüver inició una serie de investigaciones en la Universidad de Chicago, en 1926. Explicó que la mescalina inducía imágenes que podían observarse tanto con los ojos abiertos como con los ojos cerrados y que, si se mantenían los ojos abiertos, era imposible mirar a una pared blanca sin que rápidamente se viera cubierta por diversas formas.

Entre estas formas, Klüver halló cuatro tipos constantes. Uno de ellos lo describió como un enrejado, red, calado, filigrana, alveolado y tablero. El segundo tipo parecía una telaraña. El tercero lo describía en términos de túnel, corredor, cono, embudo y vasija. El cuarto tipo consistía en espirales. Las constantes de forma se caracterizaban, además, por una serie de colores variados y saturados, brillo intenso y configuraciones simétricas. Las visiones parecían hallarse a una distancia de lectura; variaban mucho en su tamaño aparente. En general, no podían controlarse conscientemente.

Klüver hizo la observación crucial de que estas constantes de forma aparecían en una amplia variedad de condiciones alucinatorias. Enumeró una serie de tales condiciones, lista que han incrementado otros investigadores, y así hoy comprende el dormirse, el despertarse, la hipoglicemia insulínica, el delirio de la fiebre, la epilepsia, los episodios psicóticos, la sífilis avanzada, la privación sensorial, la fotoestimulación, la estimulación eléctrica, la mirada vidriosa, las migrañas, los vértigos y naturalmente una amplia gama de intoxicaciones por drogas.

La mayoría de las drogas que dan lugar a dicha imaginería entran dentro de la clasificación de los alucinógenos. Sin embargo, otras drogas y sustancias pueden dar lugar a efectos parecidos; así, los compuestos psicoactivos (por

extender el ámbito a los elementos que pueden causar alteraciones de conciencia o atención) pueden considerarse como alucinógenos. Entre ellos podemos enumerar el alcohol, el dióxido de carbono, la cocaína, el cortisol, la digitalis, la escopolamina e incluso el tabaco con una fuerte concentración de nicotina.

Las constantes de forma aparecen en el primero de los dos estadios de las imágenes inducidas por las drogas. Las imágenes del segundo estadio, que son más complejas, si bien pueden incorporar las constantes simples, abarcan paisajes, caras y objetos y lugares familiares. Las imágenes complejas, tal vez el aspecto más llamativo de la experiencia alucinatoria, se consideran como una activación de las imágenes que ya se hallaban registradas en la memoria.

Podría esperarse que las formas y escenas de las imágenes complejas fueran infinitamente distintas. Realmente, las constantes aparecen incluso a este nivel. Una revisión de más de 500 alucinaciones inducidas por la dietilamida del ácido lisérgico (LSD) revelaron que mientras del 62 al 72 por ciento de los sujetos experimentaban las constantes formales simples, más del 79 por ciento expresaron haber visto imágenes complejas muy similares. Incluían símbolos e imágenes religiosas (72 por ciento) e imágenes de animales pequeños y de seres humanos (49 por ciento), la mayoría ligados afectivamente y muchos de ellos caricaturizados como personajes de historietas.

Ni los investigadores ni los sujetos describen, en general, las imágenes complejas con detalle. Además, antes del trabajo clásico de Klüver, se sabía poco sobre la geometría de las imágenes alucinatorias por sencillas que fueran. Klüver atribuyó esta omisión a la nove-



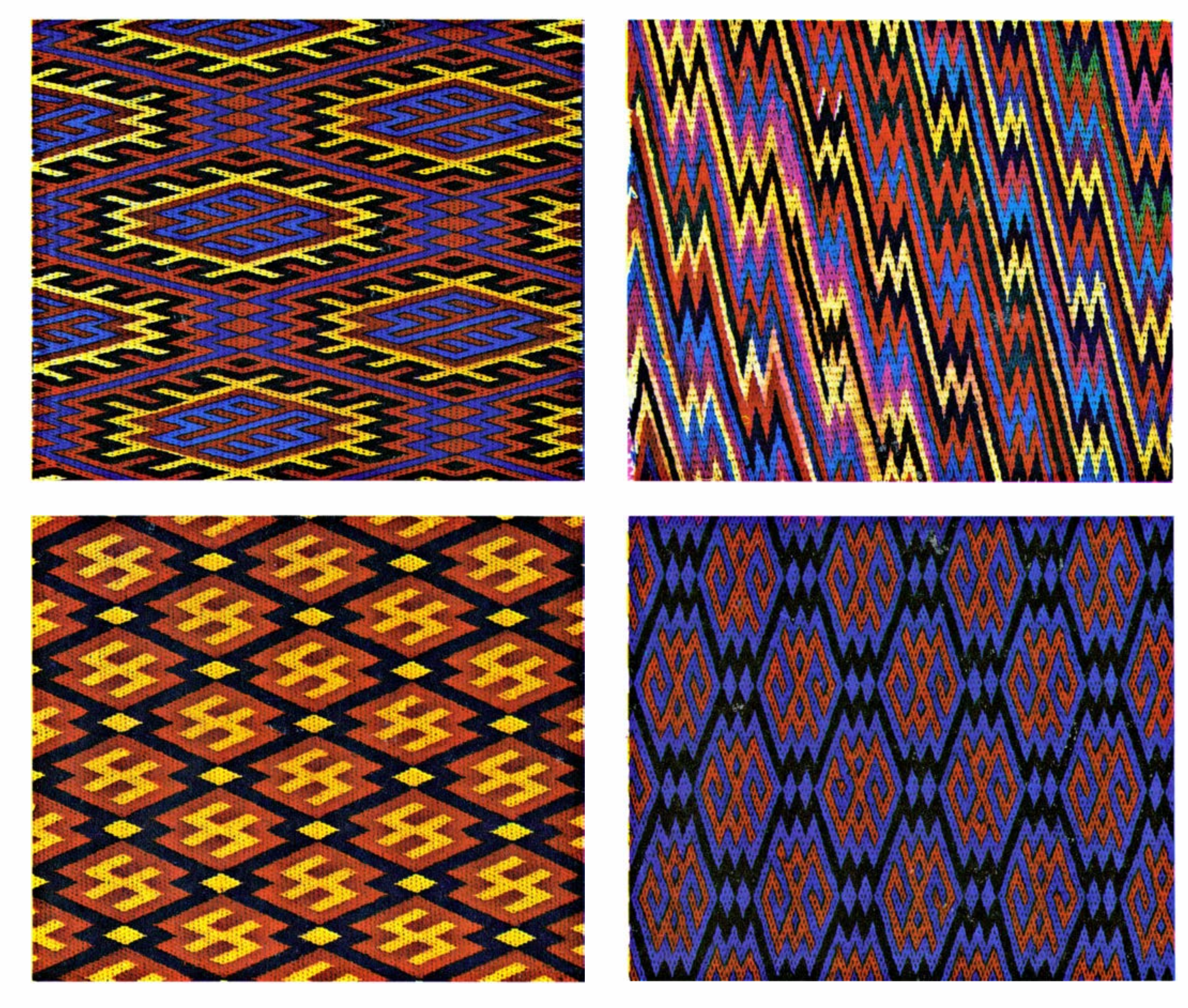
FORMAS Y COLORES ALUCINATORIOS, representados en esta pintura de un miembro de la tribu huichol de México. La pintura muestra las visiones del tipo de las que se registran en la alucinación provocada por el consumo de peyote. El indio de la izquierda lleva un cesto con peyote fresco, y padece una visión que explota en colores y destellos de luz. El cactus del peyote está representado a la derecha. Esta pintura se realiza a través de la extensión de cera en una plancha de madera para, más tarde, ir grabando las formas imaginadas sobre la cera.

dad de las visiones, expresando la idea de que muchas personas alucinadas se sienten tan atraídas por el color o el brillo que no aciertan a configurar la forma básica de las imágenes. Esta falta de articulación movió a Klüver a describir el sencillo primer estadio de las imágenes alucinatorias. La aparente complejidad de las imágenes del segundo estadio nos retó a que nosotros hiciésemos lo mismo. Creíamos que el estudio de tales fenómenos podría apuntar a una clarificación de la imaginería visual común, subyacente a las alucinaciones, que de este modo habría de ayudarnos a comprender el origen de esas percepciones y de otras relacionadas con ellas.

Nuestros primeros experimentos se diseñaron para observar si los fenómenos alucinatorios que he descrito se daban cuando las alucinaciones se inducían en sujetos no entrenados en el consumo de droga. A cada sujeto se le facilitaba o bien una dosis convencional de alucinógeno (normalmente marihuana o su principio activo, el tetrahidrocannabinol) o bien un placebo inactivo (el individuo no sabía qué sustancia recibía). Se le pedía entonces que se tumbara sobre una cama, en una habitación a salvo de la luz y el sonido, y que explicara sus experiencias. Las explicaciones se registraban en cinta magnetofónica y se las analizaba de acuerdo con la frecuencia de las distintas formas, colores, movimientos e imágenes complejas.

Los resultados demostraron que la imaginería normal (es decir, aquella que no se induce por drogas) se caracteriza por formas amorfas blancas-y-negras (a veces incluyen líneas y curvas) que se mueven de un modo más o menos aleatorio por el campo visual. Cualquiera que cierre los ojos o entre en una habitación oscura experimentará una imaginería básicamente de este tipo. Se pueden inducir imágenes similares, algunas brillantemente coloreadas y geométricas, si se fricciona con suavidad los párpados entornados.

Ahora bien, con drogas alucinógenas el número de imágenes expresadas por los sujetos se elevó muchísimo. La opinión común era que las imágenes se parecían a lo que se puede ver en una proyección cinematográfica o de transparencias. Muchos individuos tuvieron dificultades para describir sus imágenes, pero estuvieron de acuerdo en señalar que abundaban las formas geométricas. Las imágenes se caracterizaban por una luz brillante en el centro del campo



FORMAS DE RED, una de las diversas constantes de formas que se dan en las alucinaciones inducidas por droga y que se ejemplifican por medio de cuatro muestras de tejidos de los indios huichol. Estas estructuras ilustran diseños que normalmente se hallan en las visiones inducidas por el peyote. Algunas de las pinturas realizadas por esquizofrénicos exhiben una preocupación similar por los diseños geométricos.

visual, que obscurecía los detalles, si bien permitía que se observaran las imágenes periféricas.

La ubicación de este punto luminoso creaba una perspectiva similar a la de un túnel. Los individuos comentaron que veían la mayor parte de sus imágenes en relación con un túnel. Según sus narraciones, las imágenes tendían a vibrar, moverse hacia el centro del túnel o lejos de la luz brillante, y a veces se desplazaban en ambas direcciones.

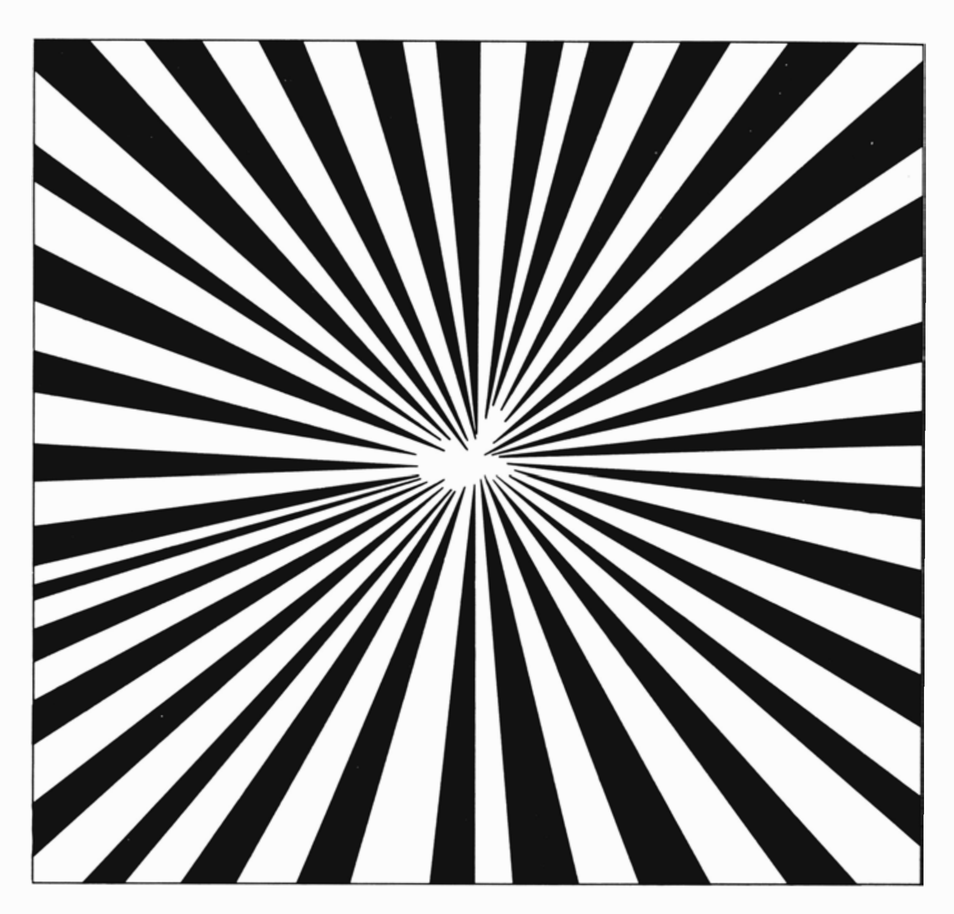
Cuando las imágenes aparecían en color se mencionaban todos los colores, aunque los informes apostillaban la incidencia de que el rojo aumentaba conforme crecía la dosis de la droga. Las formas geométricas se combinaban, duplicaban y sobreimponían frecuentemente. A veces, el flujo de imágenes era tan rápido que la mayoría de los individuos encontraban dificultades para no perder el hilo de sus comentarios.

Las formas geométricas quedan sustituidas en seguida por imágenes complejas. A decir de los individuos sujetos a experimentación, las imágenes complejas incluían escenas, personas y objetos reconocibles, la mayoría de ellos caricaturizados o en forma de dibujo animado, con cierta profundidad y simetría. Muchas veces, las imágenes se proyectaban sobre un fondo de formas geométricas.

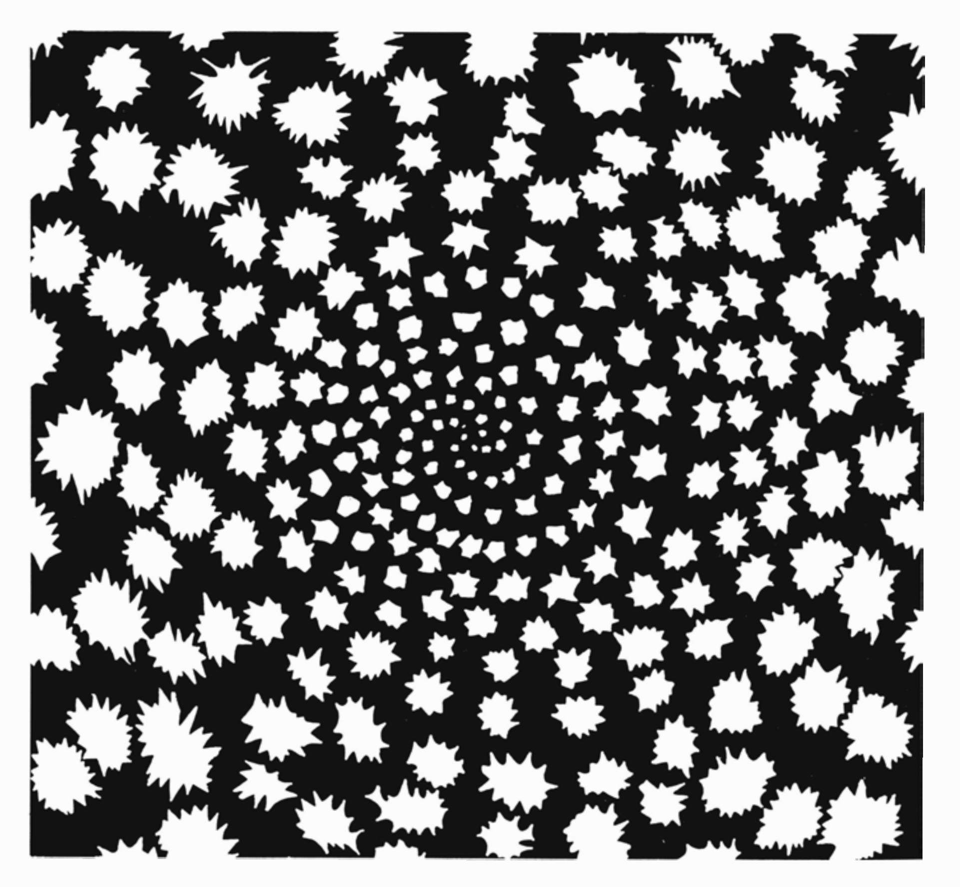
Al escuchar estos informes, topábamos con ciertas dificultades. Los individuos escogían de modo diverso las palabras para su exposición. Además, los informes estaban teñidos de experiencias idiosincrásicas. En consecuencia, optamos por facilitar la fluidez y exactitud del informe mediante el entrenamiento de nuestros sujetos en el empleo de un código de descripción convencional.

Los psicólogos Ogden R. Lindsley y Timothy Leary, de la Universidad de Harvard, habían realizado intentos anteriormente siguiendo esta línea. Emplearon la manipulación de pulsadores como medio de información: cada pulsación se hacía corresponder con la ocurrencia de un estado subjetivo. El individuo en trance de alucinación que quería informar que "veía" u "oía" algo, apretaba el pulsador indicado para cada condición. Lindsley y sus colegas habían demostrado ya que la técnica de la pulsación proporcionaba un método continuo y objetivo para el estudio del comportamiento en los estados de alteración de conciencia.

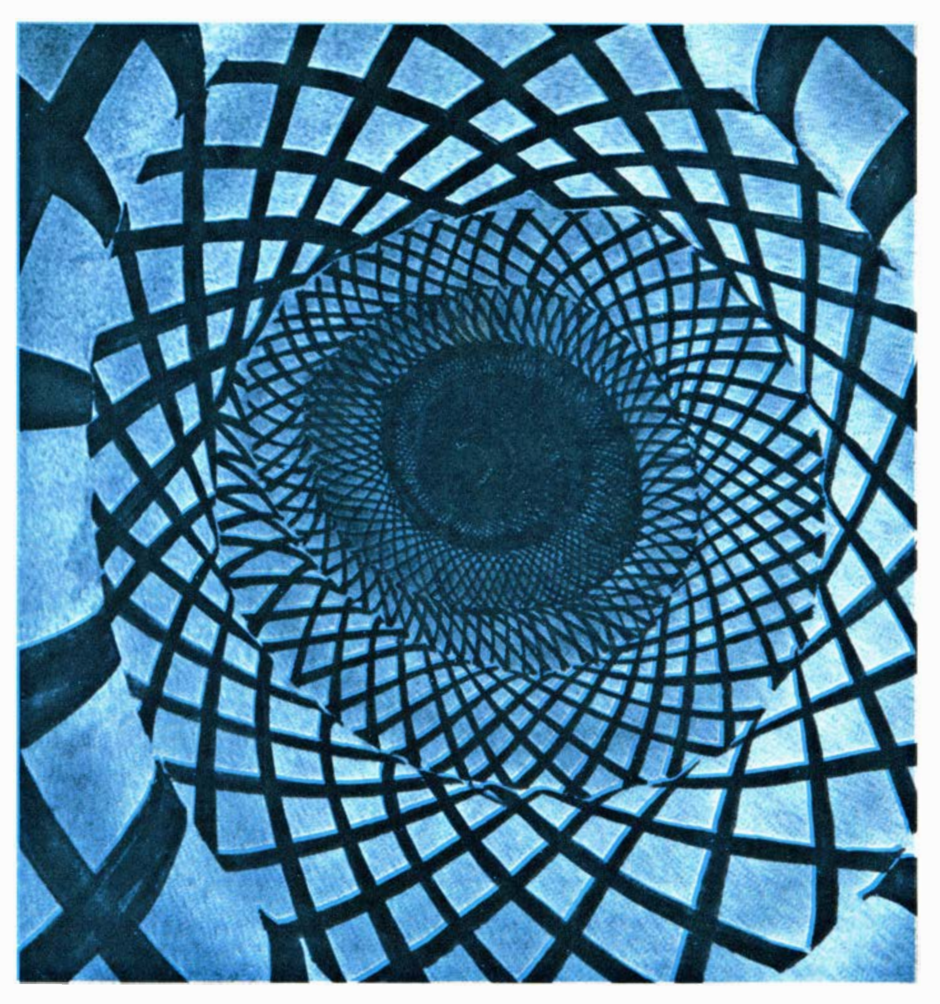
A pesar de ello, no se desarrolló un método para obtener información detallada de los fenómenos subjetivos hasta



LUZ BLANCA, vista durante las primeras etapas de intoxicación con una droga alucinógena. Las imágenes visuales se irradian desde el centro a la periferia. La estructura aparece en principio en blanco y negro, pero a medida que va avanzando la experiencia pueden aparecer colores brillantes.



TUNEL EN ESPIRAL, otra de las primeras formas constantes en las alucinaciones inducidas por droga. Como principales pautas de movimiento que le acompañan se dan la rotación y vibración.



FORMA CONSTANTE EN RED-TUNEL, expresada en una pintura realizada para mostrar una pauta que aparece normalmente en los primeros estadios de la intoxicación por marihuana y tetrahidrocannabinol (principio activo de la marihuana). Es posible que el color azul se relacione con el descenso inicial de la temperatura corporal y con la absorción de la luz azul por la hemoglobina en las células "flotantes" retinianas. Con el tiempo y más dosis, el color tiende a volverse rojo.

que Lindsley y Leary aplicaron la "máquina de escribir experiencial". Este aparato consistía en una máquina de escribir de 20 teclas conectada a un instrumento de registro. Las teclas servían para codificar distintos estados subjetivos, incluyendo modos de percepción, imágenes internas y externas, imágenes vistas con los ojos cerrados, imágenes vistas con los ojos abiertos, alucinaciones, colores, etcétera. El entreno consistía en que los sujetos memorizaran las categorías y las teclas correspondientes. Cuando un sujeto sentía una sensación corporal como de "dolor", pulsaba la tecla indicada para significar tal acontecimiento; pulsando otra tecla podía significarse una alucinación con colores oscilantes.

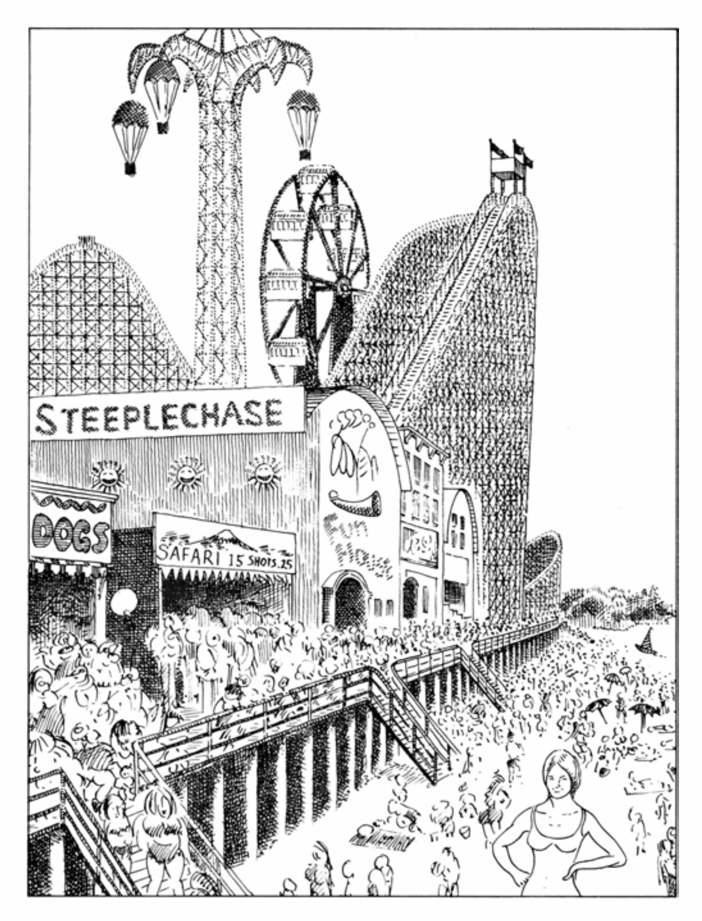
En las pruebas con los alucinógenos LSD y dimetiltriptamina, Leary halló que la incoordinación muscular asociada a tales drogas interfería el accionamiento de las teclas. El problema podría haberse evitado con dosis menores, y el informe por tecleo substituirse por un informe verbal. (Leary había observado que el informe verbal no quedaba sometido al influjo de las drogas.) Por ello es por lo que nosotros decidimos desarrollar un código verbal para informar sobre las alucinaciones inducidas por la droga.

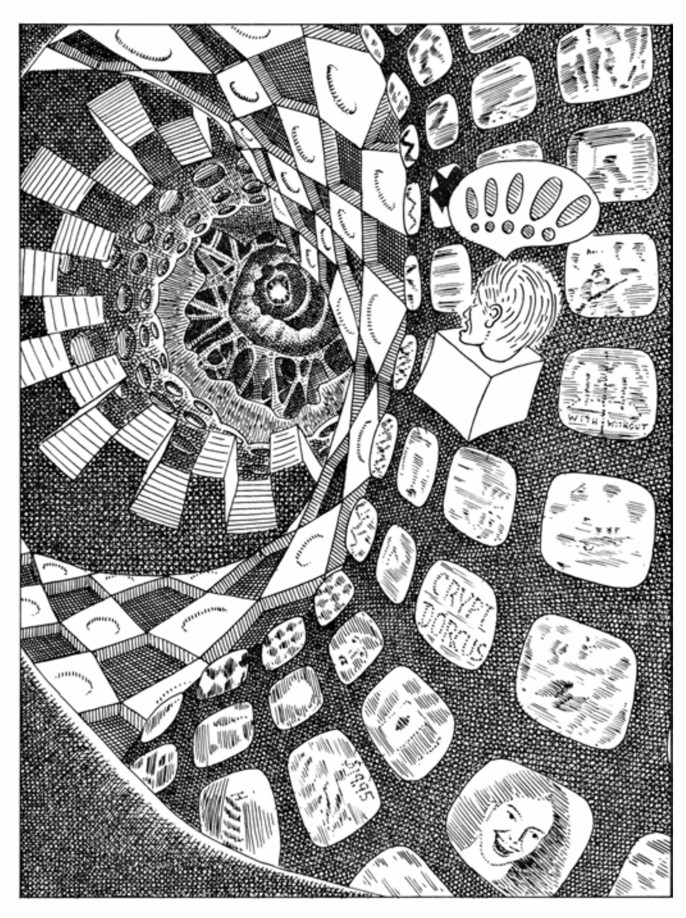
Basándonos en los resultados anteriores, elaboramos una lista de ocho formas (azar, línea, curva, trama, rejilla, túnel, espiral y calidoscopio), ocho colores (negro, violeta, azul, verde, amarillo, naranja, rojo y blanco) y ocho estructuras de movimiento (desordenado, vertical, oblicuo, horizontal, explosivo, concéntrico, rotativo y vibratorio) con el propósito de que las emplearan los individuos sujetos a experimentación para descubrir la imaginería visual. Seleccionamos luego un grupo de individuos y los entrenamos por medio de transparencias que representaban las distintas categorías. Por ejemplo, en el entreno referente a la forma de túnel, les mostrábamos cientos de transparencias diferentes de túneles, a fin de que los individuos poseyeran una idea bastante definida de lo que era un túnel. De esta manera nuevos ejemplos de formas, que quizá no se hubieran percibido antes, podían clasificarse adecuadamente en base a los rasgos comunes. Se entrenaba a los individuos en el reconocimiento de las tres categorías (forma, color y movimiento) mediante proyecciones de 8 milisegundos de duración, con una pausa de un segundo entre proyecciones. Al hacer que las proyecciones fuesen breves, pretendíamos simular los rápidos cambios de imágenes que se dan en las alucinaciones.

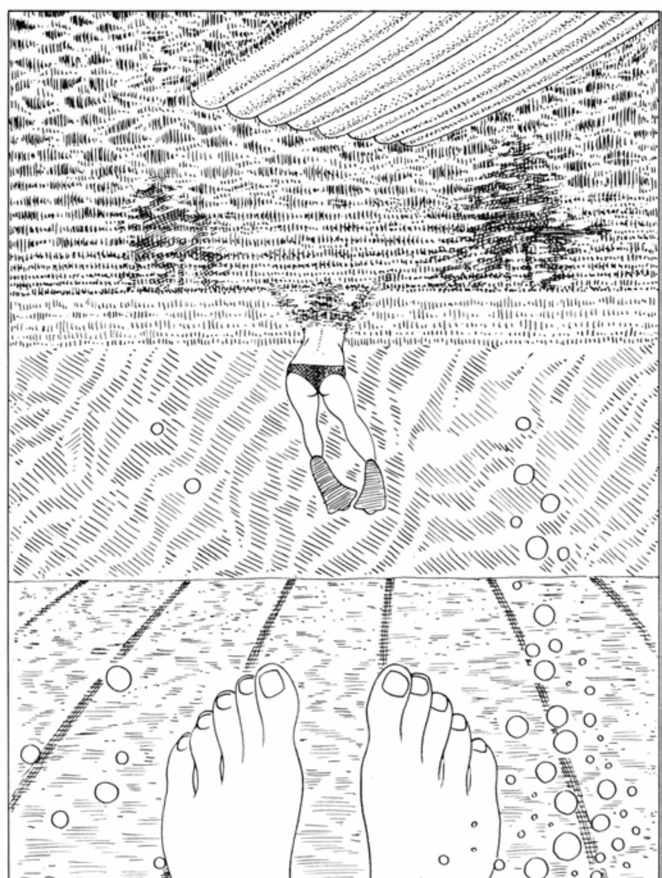
Todos los individuos, incluidos algunos que no se sometieron al entrenamiento al objeto de que se nos permitiera ver si sus informes de experiencias visuales podían compararse con los de los individuos entrenados, participaron después en una serie de sesiones de ensayos semanales, en cada una de las cuales recibían un alucinógeno, un estimulante, un depresor o un placebo. Ningún individuo sabía lo que recibía; se probó con dosis altas y con dosis bajas. Las pruebas se realizaban en la habitación libre de luz que ya hemos mencionado. Se instruía a todos los sujetos para que informasen de lo que "veían", manteniendo abiertos los ojos; a los individuos entrenados se les exigía que se expresasen mediante el código descriptivo. A mitad de la sesión, los individuos entrenados pasaban una prueba con transparencias de imágenes reales para asegurar que tanto las drogas como las dosis no hacían perder las habilidades que habían conseguido durante el entrenamiento. (No se halló ninguna señal de que ocurriera tal cosa.)

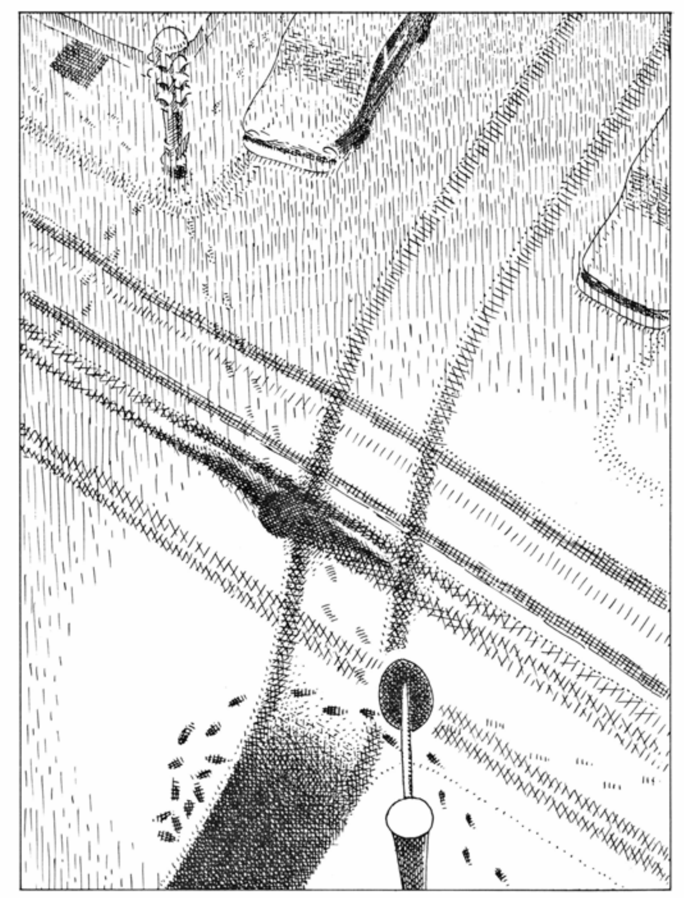
Los resultados fueron intrigantes. Hallamos que los individuos entrenados podían soportar el rápido paso de imágenes y las podían clasificar, en su mayoría, en las categorías del código de información establecido. Estos individuos realizaban un promedio de 20 informes por minuto; los individuos no entrenados daban cuenta solamente cinco veces por minuto. (La duración de una sesión normal era de 6 horas.)

Las imágenes asociadas con los placebos, el estimulante *d*-anfetamina y el depresor fenobarbital, se describían como formas aleatorias blancas-y-negras, que se movían sin orden ni concierto. Sin embargo, los alucinógenos tetrahidrocannabinol, psilocibin, LSD y mescalina inducían grandes cambios. Las formas eran aquí menos aleatorias, más organizadas y geométricas a medida que avanzaba la experiencia. Las imágenes blancas-y-negras empezaban a colorearse de azul y el movimiento tendía a ser más organizado y vibratorio. Al





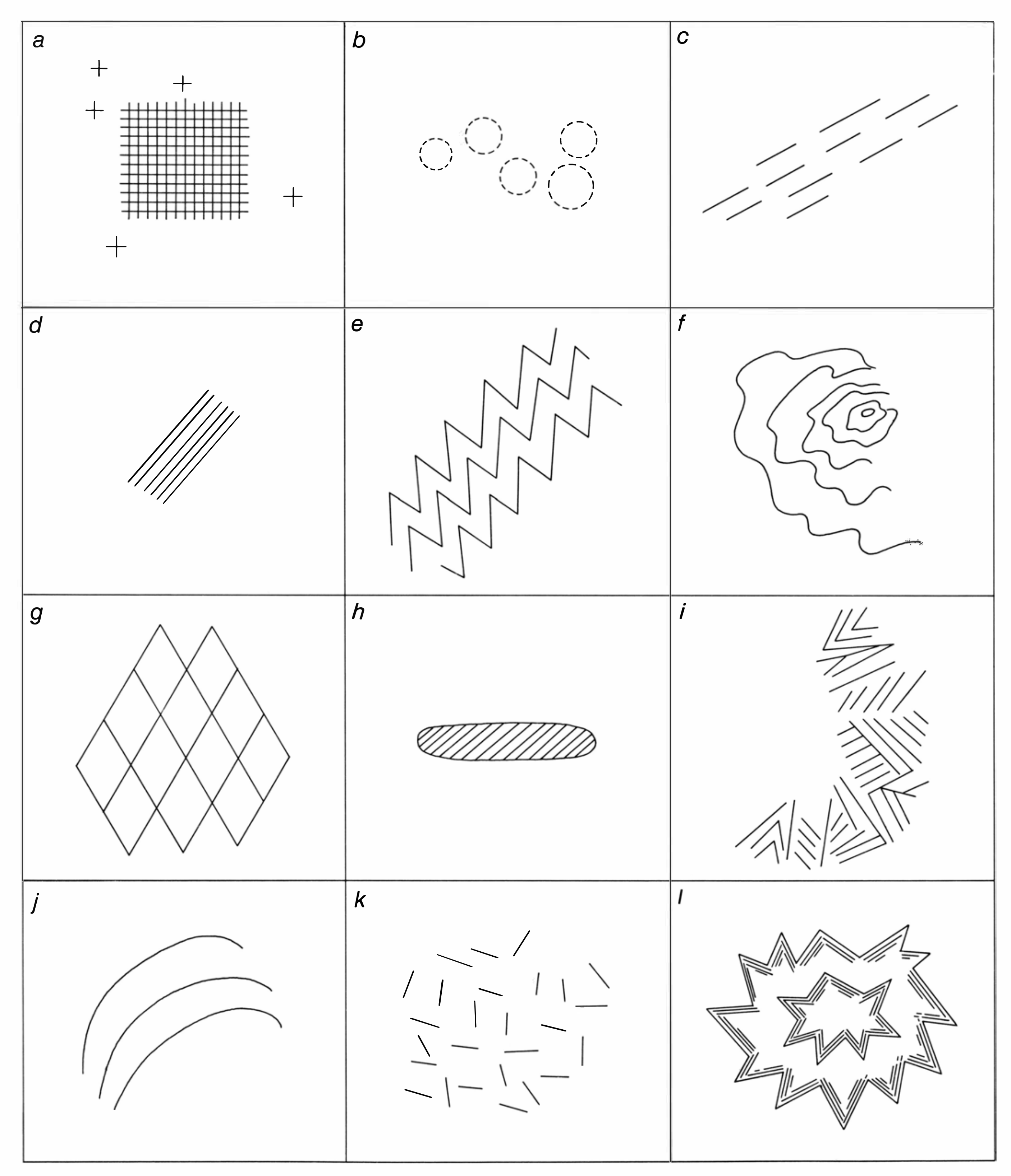




PERSPECTIVAS POCO USUALES de las que se encuentran imágenes alucinatorias que proceden, según parece, de la memoria. Las perspectivas características incluyen una escena distante (con muchos detalles), normalmente identificada con un suceso de la infancia *(arriba, a la izquierda);* una estructura de red-túnel con imágenes mnésicas complejas en la periferia *(arriba, a la derecha);* una escena vista como si el individuo estuviese bajo el agua, mirando arriba y a través la superficie *(abajo, izquierda)* y, por último, una perspectiva aérea *(abajo, a la derecha).*

cabo de 30 minutos de haberse administrado la droga, los individuos informaban de un importante aumento de las formas de túnel y red y un ligero descenso de las formas caleidoscópicas. Entre los 90 y 120 minutos, la mayoría de las formas eran túneles-redes. Concomitantemente, los colores derivaban hacia el rojo, naranja y amarillo. El movimiento continuaba siendo vibratorio, pero mucho más organizado, con estructuras explosivas y rotativas.

La imaginería compleja no solía aparecer hasta que ya se había informado de la derivación hacia las formas red-túnel. Luego, entre un 45 y un 75 por ciento de la totalidad de formas de que daban cuenta los individuos entrenados que habían recibido alucinógenos, eran de este tipo. Al principio, las imágenes complejas aparecían como cubriendo las



ALUCINACIONES VISUALES contempladas durante una intoxicación controlada de cocaína, dibujadas por los individuos del experimento ideado por el autor. Generalmente se veían las estructuras con los ojos abiertos en una habitación obscura. Aparecen como puntos blancos y negros en la periferia del campo visual. Las estructuras *c, e, f, g, i, k,* y *l* son virtualmente idénticas a las que se ven en las alucinaciones que acompañan a los ataques de migraña. La pauta *l* es la "ilusión de fortificación", que suele darse en personas que padecen ataques de migraña.

estructuras red-túnel situadas en la periferia de tales imágenes.

Las imágenes complejas más corrientes comprendían recuerdos de infancia y escenas asociadas con experiencias emocionales fuertes que habían padecido los individuos. Estas imágenes alucinatorias no eran simples réplicas gráficas; la mayoría de ellas se elaboraban y adornaban con escenas fantásticas. Este aspecto constructivo de la imaginería puede ilustrarse mediante un ejercicio sencillo. Consiste en recordar la última vez que se nadó en el mar y pensar si este recuerdo contiene una imagen de uno mismo corriendo por la playa o sumergido en el agua. Tal imagen, obviamente, es ficticia, ya que es imposible que uno pueda verse a sí mismo; pero las imágenes de nuestra memoria incluyen muchas veces recuerdos fugaces de este tipo. Nuestros individuos nos comunicaban muchas veces imágenes igualmente improbables, tales como perspectivas aéreas y visiones subacuáticas.

No ha podido establecerse de dónde surgen estas imágenes, pero fácilmente se puede pensar en la cooperación de la codificación, almacenamiento y recuperación que se da en el proceso de la memoria. Gran parte del contenido de las imágenes complejas puede estar influido por estímulos ambientales. Ahora bien, puesto que nuestros individuos se hallaban solos en una habitación aislada, la mayoría de sus imágenes provenía de su memoria. Sin embargo, de forma ocasional, llevábamos a nuestros sujetos (en el momento de mayor fuerza de las alucinaciones) a un jardín botánico. Allí se les colocaban gafas opacas que no les permitían ver, se les tendía en la hierba e informaban de lo que "veían" con los ojos abiertos. En estos casos, las imágenes memorísticas se reducían notablemente, y aumentaban los informes sobre pájaros, aviones y árboles y demás. En otros experimentos habíamos despertado y provocado en los individuos imágenes inducidas por droga comunicándoles palabras o poniéndoles fragmentos musicales sugestivos. Incluso en estos momentos, sin embargo, los individuos expresaban que las imágenes complejas aparecían normalmente en combinaciones con redes-túneles y moviéndose en configuraciones explosivas o rotatorias.

En el punto álgido de los períodos alucinatorios, los individuos se describían frecuentemente a sí mismos como parte integrante de sus imágenes. En tales ocasiones dejaban de dar símiles en sus informes y aseguraban que sus imágenes eran reales. Este punto señalaba el paso de la falsa a la verdadera alucinación. Hablaban entonces de combinaciones fantásticas y altamente creativas, con una velocidad de unas 10 imágenes por segundo. Los sujetos solían contar que se sentían disociados de sus cuerpos.

Parece como varios remolinos diferentes, con muchas espirales divididas en cuadrados. Es bastante negro. Hay zonas púrpura y verde, fulgurantes, en el medio de las espirales, como nubes alrededor. Hay líneas que van de arriba abajo, como una rejilla, pero las líneas se retuercen en torno. Hay formas raras, pero aún hay cantidad de ángulos rectos en ellas. Parece realmente brillante... Hay como una explosión, amarilla en el centro, como un volcán vomitando lava, amarilla, brillante. Hay un cuadrado negro con una luz amarilla que sale de detrás. Hay un patrón superpuesto a todo, muchos rizos, con lunares en el medio. Muchas cositas tejidas que llenan los espacios en medio de los patrones de triángulos, de cuadrados o de cosas en forma de corona. Y hay una estrellita blanca que flota en torno al cuadro y a veces va a ocultarse detrás de lo que está en la pantalla y lo ilumina por detrás.

Ahora hay una especie de paisaje. Terreno llano, muy llano. La imagen es muy angosta. En su parte media un árbol a la izquierda y luego llano con hierba verde y cielo azul arriba. Hay lunares naranjas, naranjas colgando por doquier, en el cielo, sobre el árbol, en el suelo. ¡Una bicicleta! ¡Caramba! Se dirige hacia abajo, no horizontal, como si alguien la sujetara de un extremo... Hay un tablero de ajedrez superpuesto a todo, como las banderas que se agitan en las carreras.

Puedo ver la calle allí fuera... Bueno, es antigua –¡caray!– ¡interesante! Es como de los años cuarenta, supongo, o quizá los cincuenta... Y hay gente montando en sus bicicletas, y hay, algo como, muchachos, con chalecos de tela escocesa y esos sombreros tan raros... Yo iba al lado caminando sobre la acera, de modo que no era como si estuviera en el medio de la calle y (risa) uno no puede reír largo rato en el medio de la calle en la ciudad, así que la imagen se desvaneció (risa).

TRANSCRIPCION PARCIAL del informe de un individuo que había tomado 20 miligramos de psilocibin alucinógeno y que muestra una complejidad creciente de imágenes. El primer párrafo procede de un segmento inicial de la experiencia, que empezó al cabo de 25 minutos de que el individuo hubiese tomado la droga. Las formas geométricas simples se complementan varios minutos después *(segundo párrafo)* por imágenes complejas. Posteriormente *(párrafo tercero)*, conforme avance el tiempo del ensayo, se verán substituidas por escenas similares a las de los sueños.

Las constancias notables de las alucinaciones inducidas por droga llevan de forma natural al planteamiento de la cuestión de su universalidad. Algunas resultan sorprendentemente semejantes a las formas arquetípicas o primordiales (como son el mandala, el símbolo místico del universo empleado por el Brahamismo y el Budismo como ayuda para la meditación) y que el psicoanalista C. J. Jung describió como una parte del inconsciente colectivo del hombre. Además, tal como han señalado muchos antropólogos, el arte inspirado en los alucinógenos de muchos pueblos primitivos contiene constantes de forma, color y movimiento.

Examinamos este fenómeno en una expedición a la Sierra Madre de México, para estudiar un grupo de indios huichol, consumidores de peyote. Se habían mantenido relativamente aislados desde los tiempos de los aztecas. Les entrevistamos durante las ceremonias en las que consumían peyote, elicitando respuestas sobre su imaginería visual. Se comprobó que las imágenes eran virtualmente idénticas a las estructuras simétricas y repetitivas que caracterizan a las formas de expresión artística de los huichol.

Al intentar explicar el origen de las constantes alucinatorias simples, son muchos los investigadores que las han descrito como producto de fenómenos intraoculares resultantes de la visualización de ciertas estructuras cuando la luz penetra en el ojo de una forma determinada. Por ejemplo, se dice que las bandas horizontales se deben a la estimulación del epitelio corneal, que cambia con el movimiento palpebral, y se atribuye la visión de un encaje negro sobre un fondo rojo a un aumento de irrigación de los conos y bastones por parte de los vasos sanguíneos de la retina.

Ahora bien, puesto que estas causas necesitan de la luz, no pueden haber desempeñado un papel importante en nuestros experimentos, que se realizaban en la oscuridad. La luz no es necesaria para la producción de fosfenos, que son las sensaciones visuales procedentes de la descarga de neuronas en estructuras del ojo. Los fosfenos pueden abarcar puntos, discos, arcos o círculos concéntricos y configuraciones en tablero de ajedrez.

Las constantes son también muy similares a las estructuras que pueden hallarse en las alucinaciones que acompañan a los ataques de migraña. Las pautas de las migrañas comprenden líneas, enrejados, círculos concéntricos y la "ilusión de fortificación", que consiste en una zona en forma de herradura con líneas brillantes en zigzag, que parecen expandirse desde el exterior. Las alucinaciones de la migraña pueden estar también brillantemente coloreadas y ser explosivas; asimismo pueden incluir imágenes complejas de personas y objetos. La explicación más plausible para los fenómenos de la migraña es que reflejan la excitación eléctrica de grupos organizados de células, del cortex visual del cerebro. Además, el trabajo de Wilder Penfield y sus colegas, de la McGill University, ha demostrado que la estimulación eléctrica directa del cortex visual o de los lóbulos temporales da lugar a luces coloreadas, formas geométricas, estrellas y líneas en movimiento.

La mayoría de los investigadores que intentan explicar la imaginería alucinatoria compleja, describen las imágenes como el resultado de una excitación del sistema nervioso central. Ya en 1845 Moreau mantenía que las alucinaciones eran fruto de una excitación cerebral, que permitía que los recuerdos y pensamientos se transformaran en impresiones sensoriales. La investigación electrofisiológica más reciente ha confirmado que las alucinaciones se relacionan directamente con estados de excitación y vigilia del sistema nervioso central, que van unidos a una desorganización de la parte del cerebro que regula la llegada de estímulos exteriores. Desde el punto de vista del comportamiento, el resultado es un deterioro de la discriminación que se basa normalmente en los estímulos externos y una preocupación por las imágenes interiores.

Se ha definido al proceso alucinatorio como una escena retrospectiva ("flashback") de la memoria y una "reminiscencia involuntaria". Ciertos psicoanalistas han argumentado que se trata del resultado de una regresión a un estado similar al pensamiento infantil, conjuntamente con la aparición de información y recuerdos reprimidos. Los estudiosos de los fenómenos psicodélicos sostienen que los alucinógenos dejan libres informaciones y recuerdos refrenados.

Posiblemente, la explicación más completa se contenga en la teoría de la liberación perceptiva, formulada por el neurólogo británico Hughlings Jackson, en 1931. Tal como recientemente lo señaló Louis Jolyon West, de UCLA, la hipótesis mantiene que los recuerdos normales se refrenan por medio de un mecanismo que actúa como una puerta del flujo de información que proviene del exterior. Una entrada de información nueva inhibe la aparición y conciencia de las percepciones anteriores y de la información procesada. Si esta entrada decrece o se debilita en tanto que permanece la conciencia, se consigue que estas percepciones se liberen y se organicen dinámicamente, experimentándose como alucinaciones, sueños o fantasías.

West recurrió a una analogía para ilustrar el proceso. Representémonos a un hombre en la sala de estar, que se halla junto a una ventana cerrada, opuesta a la chimenea y mirando a través de ella la puesta del sol. Está absorto en la contemplación del mundo exterior y no visualiza el interior de la habitación. A medida que va obscureciendo fuera, las imágenes de los objetos de la habitación que están a sus espaldas pueden verse débilmente reflejados en la ventana. Cuanto más cerrada se va haciendo la obscuridad del exterior, con más intensidad ilumina la hoguera el interior de la habitación, y el individuo ve ahora un reflejo vivo en la estancia, que parece exterior a la misma. Cuando la analogía se aplica a la hipótesis de la liberación perceptiva, la luz del sol (entrada sensorial) se reduce, mientras la iluminación interior (nivel general de alerta del sistema nervioso central) se mantiene brillante, de forma que las imágenes originadas en las estancias del cerebro pueden percibirse como si vinieran de las ventanas exteriores de los sentidos.

A través de estas investigaciones e hipótesis hemos empezado a comprender la naturaleza de las alucinaciones como imágenes almacenadas en el cerebro. Al igual que un espejismo que muestra una magnífica ciudad, las imágenes de las alucinaciones son realmente imágenes reflejadas de objetos auténticos ubicados en otra parte. La ciudad no vale menos ni resulta carente de interés porque no esté donde pensamos que está. El tiempo nos ayudará a localizarla.

Juegos matemáticos

De cómo jugar a la "Nueva Eleusis", pasatiempo que simula muy bien la búsqueda de la verdad

Martin Gardner

Consideraré siempre mejor profeta al autor de la mejor conjetura.
Cicerón. *De Divinatione*

Nunca profetices - a menos que sepas.
James Russell Lowell,
The Biglow Papers

Tuve el privilegio de presentar en la revista *Scientific American,* correspondiente a junio de 1959, un notable juego de simulación llamado "Eleusis". El juego, que se realiza con baraja francesa de 52 cartas, recibe su nombre de los antiguos misterios eleusinos, ritos religiosos en los que los iniciados aprendían los arcanos de un culto secreto. Se han preparado cientos de ingeniosos juegos de simulación con la finalidad de obtener modelos de distintos aspectos de la vida. A pesar de ello, Eleusis es particularmente interesante para matemáticos y científicos porque proporciona un modelo de proceso de inducción, proceso que es el corazón mismo del método científico. Mi primer artículo sobre "Eleusis" volvió a publicarse en *The 2nd Scientific American Book of Mathematical Puzzles & Diversions* (Simon and Schuster, 1961). Desde aquella fecha, la Eleusis se ha convertido en un juego mucho más apasionante que su versión primitiva, y me siento en la obligación de actualizarlo para mis lectores. Comenzaré haciendo un poco de historia.

El juego de Eleusis fue inventado en 1956 por Robert Abbott, de Nueva York, por aquellas fechas alumno de la Universidad de Colorado. Abbott había estado estudiando lo que algunos psicólogos llaman "reacción ¡Ajá!", esa súbita comprensión profunda de un problema que conduce a su solución. Más de una vez ha cambiado de rumbo la ciencia por uno de estos misteriosos golpes de intuición. El juego de Eleusis proporciona una fascinante forma de simulación de estas facetas de la ciencia, aunque Abbott no lo inventó con esta finalidad. En 1963, Abbott dio las reglas completas del juego en su libro *Abbott's New Card Games* (encuadernado, Stein & Day; en rústica, Funk & Wagnalls).

Un eminente físico-matemático de la Princeton University, Martin D. Kruskal, se interesó en el juego y aportó varias mejoras de importancia. Sus reglas aparecieron publicadas en una monografía titulada *Delphi: A Game of Inductive Reasoning*. Muchos profesores de todo el país utilizaron la Eleusis para explicar a sus alumnos el método científico o para mostrar un modelo de "proceso Ajá". Los especialistas en inteligencia artificial pusieron a punto programas de ordenador del juego. Se hicieron trabajos de investigación sobre la Eleusis, dirigidos por J. Robert Newman, en la System Development Corporation de Santa Mónica. Litton Industries, dedicada a toda clase de rompecabezas y pasatiempos, anunció a toda plana el juego de Eleusis. Aparecieron descripciones del juego en libros y revistas europeos. Abbott comenzó a recibir cartas procedentes de todas partes con sugerencias para enriquecer su juego.

En 1973, Abbott analizó el juego con John Jaworski, joven matemático británico que había trabajado en una versión computarizada de Eleusis con la finalidad de enseñar el proceso de inducción. A raíz de ello, Abbott se embarcó en un programa de tres años al objeto de remodelar el juego, incluyendo en él tantas buenas sugerencias como fuera posible. El nuevo juego no solamente es aún más interesante; también ha ganado en riqueza metafórica. Al introducir los papeles de Profeta y Falso Profeta el juego permite ahora simular cualquier tipo de verdad. Presentaremos ahora, basándonos en un escrito de Abbott, las reglas de la "Nueva Eleusis" en la forma que actualmente la juegan los aficionados.

Se requieren al menos cuatro jugadores, pudiendo intervenir hasta ocho. Por encima de este número el juego resulta sumamente largo y se vuelve caótico.

Se utilizan dos barajas francesas completas, mezcladas conjuntamente. En ocasiones, una mano puede alargarse mucho y hacer falta un tercer mazo de naipes. Un juego completo está formado por una o varias rondas (manos de juego): en cada mano da las cartas un jugador distinto. Este jugador puede recibir nombres tales como Dios, Naturaleza, Tao, Brahma, Oráculo (así se llama en el juego de "Delfos"), o sencillamente, Repartidor.

La primera tarea del repartidor consiste en establecer una "norma secreta". Se trata, simplemente, de una regla que especifica las cartas que cada participante puede jugar lícitamente llegado su turno. Para puntuar, los jugadores deben adivinar en qué consiste la regla. Cuanto menos se tarde en descubrirla tanto mayor será la puntuación que se obtenga.

Una de las características más sagazmente elaboradas de la Eleusis es el sistema de puntuación (que se describe más adelante), diseñado de manera que al repartidor le convenga inventar una regla ni demasiado fácil ni demasiado difícil de hallar. Sin esta condición, los repartidores tenderían a imponer normas complejas que nadie podría descubrir, haciendo el juego aburrido y exasperante.

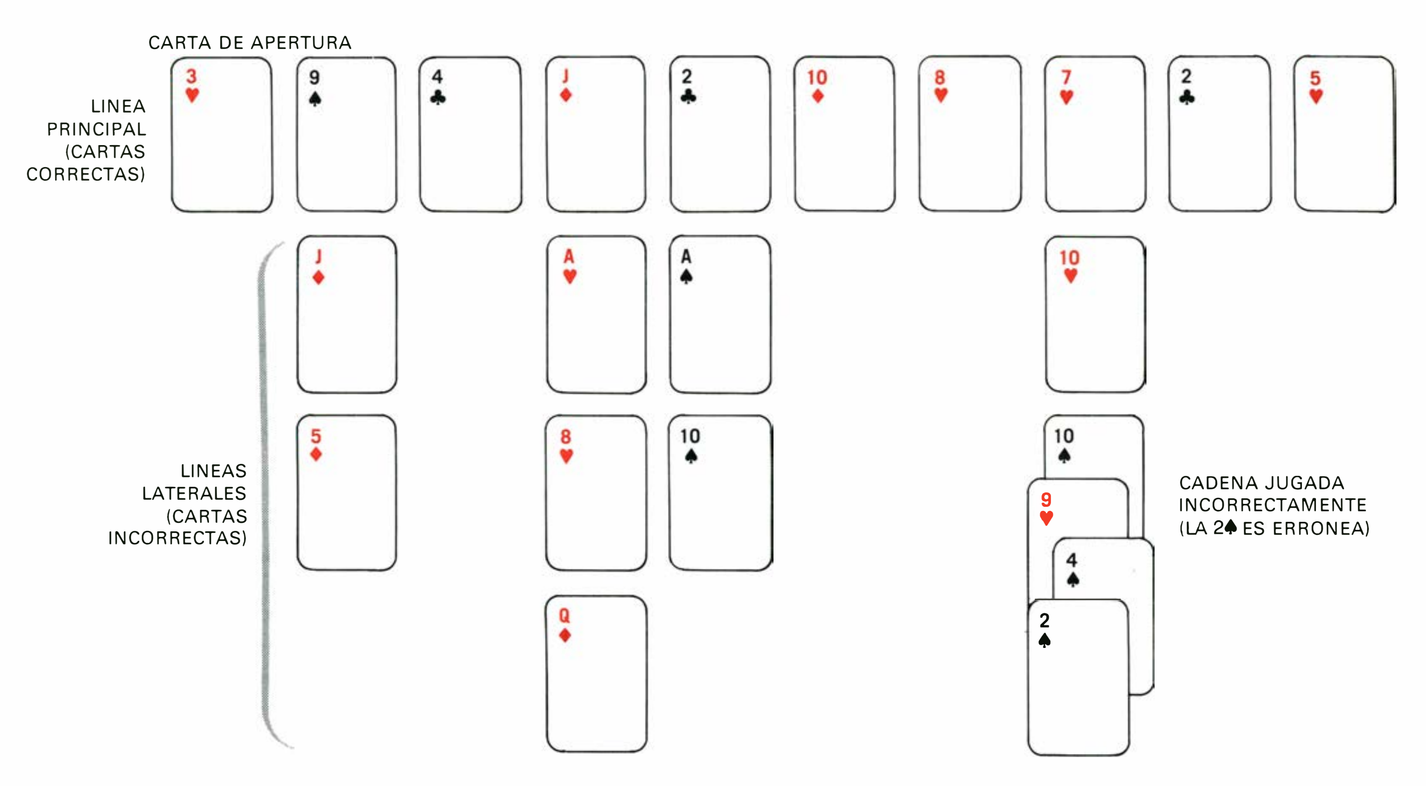
Un ejemplo de regla excesivamente sencilla podría ser: "Jugar una carta de distinto color al de la última que se haya jugado". Inmediatamente se detectaría la alternancia de colores. Algo mejor sería: "Jugar de manera que se vayan alternando cartas de valor número primo y de valor número compuesto". Sin embargo, entre matemáticos esta regla resultaría demasiado sencilla, y, para los demás, demasiado difícil. He aquí una regla excesivamente complicada: "Multiplicar los valores de las tres últimas cartas jugadas, y dividir por cuatro. Si se obtiene resto 0, jugar una carta roja o una carta de valor mayor que 6. Si el resto es 1, jugar una carta negra, o una figura. Si el resto es 2, jugar una carta par o de valor menor que 6. Si el resto es 3, jugar una carta de valor impar o un 10". Naturalmente, nadie descubriría una regla semejante, y el repartidor tendría baja puntuación.

He aquí tres ejemplos de reglas adecuadas para juegos con participantes inexpertos:

1. Si la última carta lícitamente jugada fue impar, jugar carta negra. En caso contrario, jugar carta roja.

2. Si la última carta lícitamente jugada fue negra, jugar una carta de valor mayor o igual. Si fue roja, jugar una carta de valor menor o igual. (Los valores de la sota, reina, rey y as son, respectivamente, 11, 12, 13 y 1.)

3. La carta jugada debe ser del mismo palo o del mismo valor que la última carta lícitamente jugada.



Una ronda típica de Eleusis en fase temprana.

La norma oculta debe referirse solamente a la sucesión de cartas lícitamente jugadas. Evidentemente, los jugadores muy expertos pueden utilizar normas que hagan intervenir toda la serie de cartas, lícitas e ilícitas, de la mesa, pero tales reglas son mucho más difíciles de adivinar, por lo que no están permitidas en el juego ordinario. En ningún caso puede depender la regla de circunstancias externas a las cartas. Entre los ejemplos de reglas inaceptables están las que dependen de que el jugador sea hombre o mujer, la hora del día, de si el Oráculo se rasca o no la oreja, y otras semejantes.

La regla secreta se anota en una hoja de papel, que se guarda para futura consulta y referencia, y debe estar redactada con precisión y sin ambigüedad. Como propone Kruskal, el repartidor puede dar una indicación verídica antes de que comience el juego. Por ejemplo, puede declarar que en la regla no importa el palo, o bien, por ejemplo, que la regla depende de las dos últimas cartas que se hayan jugado.

Después de anotar la regla secreta, el repartidor baraja los dos mazos de naipes y reparte 14 cartas a cada jugador y ninguna para sí. A continuación separa una carta más, llamada "carta de apertura", que se coloca en la parte izquierda más extrema de la superficie de juego, como se indica en la figura de esta misma página. Para decidir quién comienza el juego, el repartidor cuenta en sentido horario alrededor del círculo de jugadores, empezando por el situado inmediatamente a su izquierda y excluyéndose a sí mismo. Cuenta hasta llegar al valor de la carta. El jugador indicado por tal número comienza el juego, que prosigue después en sentido horario alrededor del círculo.

Una jugada consiste en colocar sobre la mesa una o más cartas. Para jugar una sola carta el jugador elige una de las que tiene en la mano y la muestra a todos. Si de acuerdo con la regla dicha carta es lícita, el repartidor dice "Buena". La carta se coloca a la derecha de la carta de apertura, en la "línea principal" de cartas correctamente jugadas, que se extiende horizontalmente hacia la derecha.

Si la carta no satisface la regla, el repartidor la declara "Mala". En este caso, la carta se coloca directamente debajo de la última carta que se haya jugado. Las líneas verticales de cartas incorrectamente jugadas se llaman "líneas laterales". (La disposición general de las cartas y la terminología se deben a Kruskal.) Así pues, varias jugadas incorrectas consecutivas prolongarán hacia bajo una misma línea lateral. Cuando un jugador juega equivocadamente una carta, el repartidor le da dos más de penalización, aumentando, por tanto, su mano.

Cuando un jugador cree haber descubierto la regla secreta puede jugar una "cadena" de 2, 3 o 4 cartas de una vez. Para jugar una cadena se superponen ligeramente los naipes a fin de respetar su orden, y se muestran a todos los jugadores. Si todas las cartas de la cadena se ajustan a la regla, el repartidor las declara "Buenas". Luego se colocan todas las cartas en la línea principal, sin solaparse, como si fueran cartas correctamente jugadas de una en una.

Si la cadena contiene una o más cartas erróneas, el repartidor declara "Mala" a toda la cadena, sin indicar cuáles son las cartas que no se ajustan a la condición. La serie errónea se deja, con solapamiento para mostrar que se trataba de una cadena, debajo de la última carta que se haya jugado. El jugador recibe el doble de cartas que tuviera la cadena.

La disposición de la figura de esta página pone de manifiesto todas las reglas de Eleusis dadas hasta ahora. La norma secreta del repartidor correspondiente a esta disposición es la primera de las tres mencionadas como ejemplo anteriormente.

Los jugadores mejoran su puntuación librándose de tantas cartas como les sea posible, lo que evidentemente se consigue de forma óptima si se descubre cuál es la regla secreta. Al comenzar cada ronda hay poca información para avanzar, y se juega necesariamente al azar. Conforme avanza la mano se va añadiendo más y más información a la disposición de cartas de la mesa, y la regla es cada vez más fácil de adivinar.

Puede ocurrir que un jugador crea conocer la regla y no disponga de cartas que jugar lícitamente. Tiene entonces la opción de declararse "sin juego". En este caso ha de mostrar sus cartas a todos. Si el repartidor lo declara correcto y si su mano no consta de más de cuatro cartas, se devuelven todas sus cartas al mazo y la ronda termina. Si el jugador está en lo cierto y tiene cinco cartas o más, se devuelven sus cartas al mazo, y se le reparte una nueva mano con cuatro cartas menos de las que tenía.

Si el jugador estaba equivocado al declarar que no tenía juego, el repartidor toma una de sus cartas correctas y la pone en la línea principal. El jugador conserva sus restantes cartas, y como penalización recibe cinco cartas más. El jugador que crea no disponer de juego correcto y no haya tampoco dado con la regla secreta debería comprender que la probabilidad de usar con éxito la opción "sin juego" no le es favorable. Le es preferible jugar al azar una sola carta.

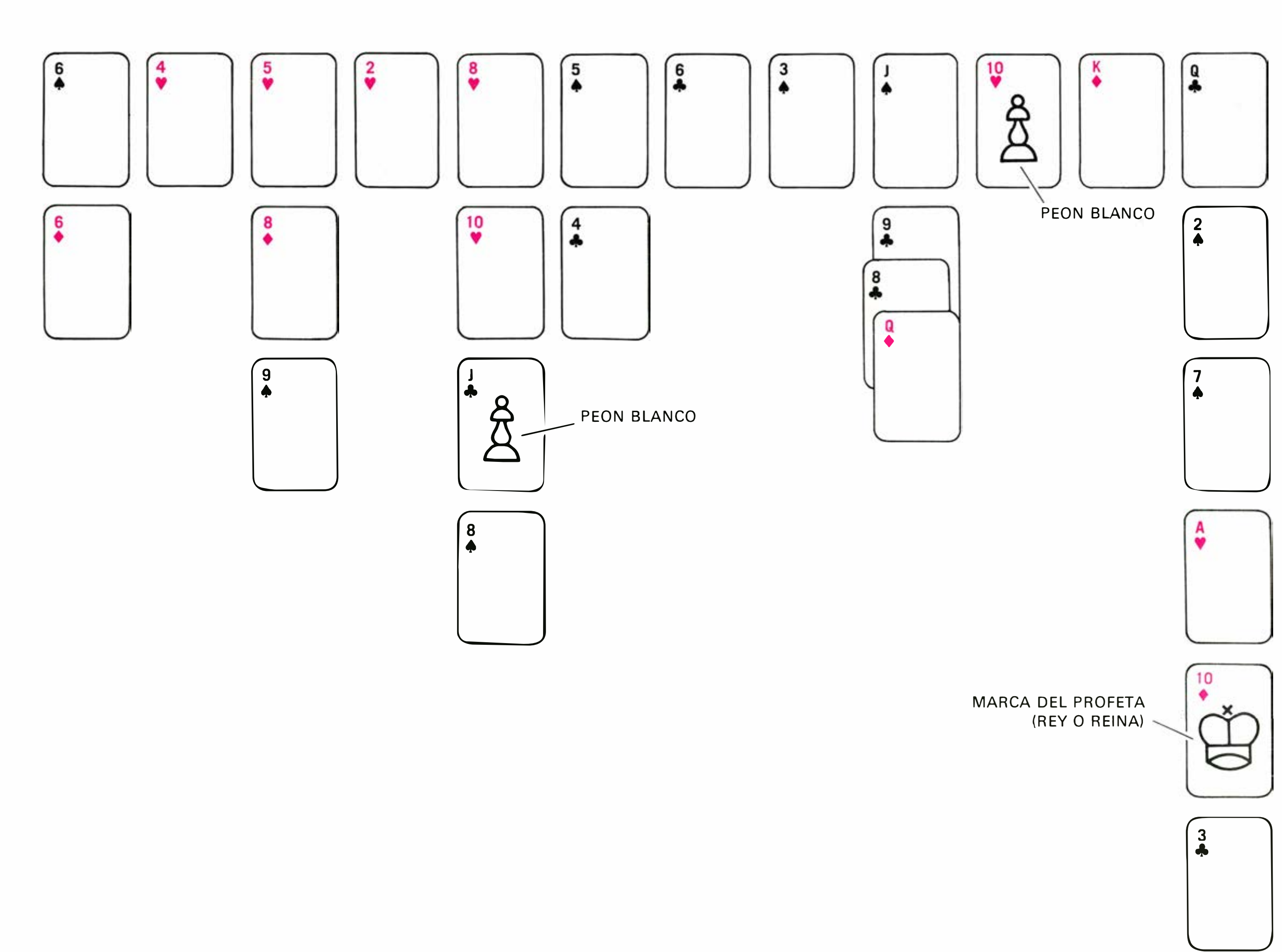
Cuando un jugador cree conocer la regla secreta tiene oportunidad de demostrar que es así y aumentar su puntuación. Para ello se declara Profeta. El Profeta asume inmediatamente los deberes del repartidor, declarando buenas o malas las jugadas de los demás y distribuyendo cartas de penalización cuando corresponda. Para declararse Profeta se requieren todas las condiciones siguientes:

1. Acabar de realizar (correcta o incorrectamente) su jugada, sin que haya jugado todavía el jugador siguiente.
2. No haber ya otro Profeta.
3. Quedar todavía en el juego al menos dos jugadores, aparte de sí mismo y del repartidor.
4. No haber sido Profeta anteriormente en esa misma ronda.

Cuando un jugador se declara Profeta pone un marcador en la última carta que ha jugado. Puede usarse un rey o reina de ajedrez. El Profeta conserva sus cartas, pero no juega ninguna hasta haber sido derrocado. El juego prosigue en sentido de las agujas del reloj alrededor del círculo de jugadores, saltándose el turno del Profeta.

Siempre que un jugador juega, sea una carta o una cadena, el Profeta declara buena o mala la jugada. Entonces el repartidor confirma o infirma la profecía, declarándola "Correcta" o "Incorrecta". Si el Profeta está en lo cierto, la carta o la cadena se sitúa en su correspondiente lugar de la mesa –en la línea principal si es buena, en una lateral si es mala– y el Profeta distribuye las cartas de penalización que proceda a los jugadores.

Si el repartidor declara "Incorrecta" la profecía, el Profeta es derrocado instantáneamente, siendo declarado Falso Profeta. El repartidor quita la marca del Falso Profeta y le adjudica cinco cartas más que añadir a su mano. No se le permite a este jugador volver a ser Profeta en la misma ronda, aunque sí puede serlo cualquier otro jugador. El simbolismo religioso es obvio, pero como señala Abbott, también hay una divertida ana-



La disposición final de una ronda de Eleusis está formada por una línea principal, varias colaterales y marcadores de diversos tipos.

logía con el mundo científico: "El Profeta es el científico que publica. El Falso Profeta es el científico que publica prematuramente". Una de las interesantes características del juego de "Nueva Eleusis" es precisamente el placer de llegar a Profeta o de expulsar al Falso Profeta.

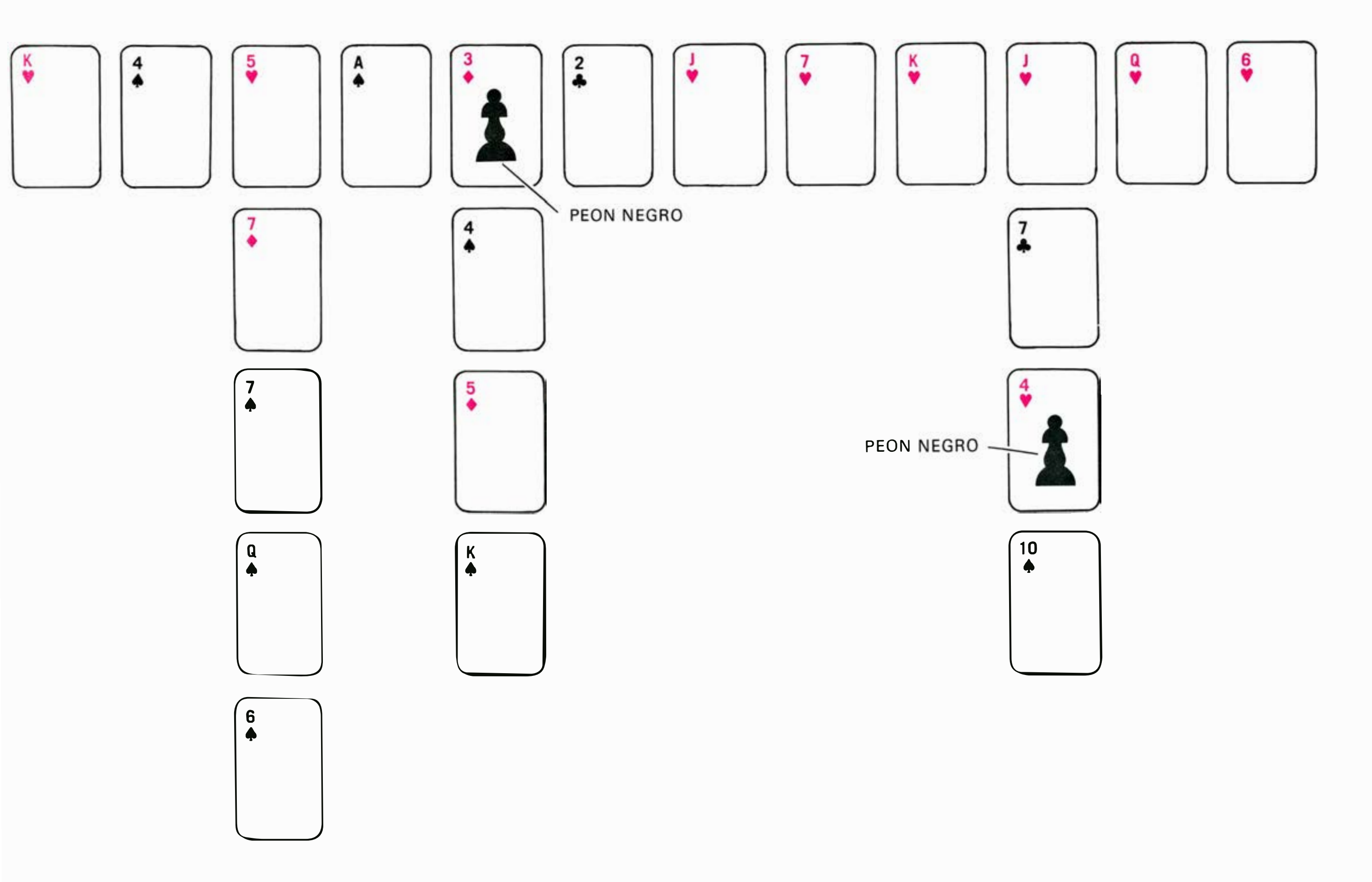
Tras la caída de un Profeta, el repartidor reasume sus primitivos deberes. Empieza completando la jugada que hizo caer al Profeta, colocando la carta o la cadena en su lugar correspondiente de la mesa. Sin embargo, si la jugada es mala no da cartas de penalización. El objetivo de tal exención es animar a los jugadores a jugadas insólitas –incluso deliberadamente erróneas– con la esperanza de derrocar al Profeta. Con terminología de Karl Popper, se trata de estimular a los científicos a pensar en métodos de "falsear" (demostrar la falsedad) las teorías dudosas de sus colegas.

Si hay Profeta y un jugador cree no tener cartas que jugar, las cosas se complican un poco. Raramente se presenta esta situación, por lo que pueden saltarse por el momento esta parte de las reglas, consultándolas cuando sea necesario. Una vez que el jugador se declara sin juego hay cuatro posibilidades:

1. El Profeta dice "Buena", el repartidor dice "Correcto". El Profeta aplica el procedimiento anteriormente expuesto.
2. El Profeta dice "Buena", el repartidor dice "Incorrecto". El Profeta es derrocado inmediatamente. El repartidor vuelve a dirigir el juego como de costumbre, aunque sin penalizar por su error al jugador.
3. El Profeta dice "Mala", el repartidor dice "Incorrecto". Dicho de otra forma, el jugador tiene razón. El Profeta es derrocado, y el repartidor vuelve a sus funciones ordinarias.
4. El Profeta dice "Mala", el repartidor dice "Correcto". En este caso el Profeta debe ahora elegir una carta correcta de la mano del jugador y colocarla en la línea principal. Si el Profeta hace correctamente la elección, penaliza al jugador con cinco cartas y el juego prosigue. Es posible, sin embargo, que el Profeta cometa un error en este punto y tome una carta incorrecta. Si tal ocurre el Profeta es derrocado. La carta errónea vuelve a la mano del jugador, y el repartidor vuelve al procedimiento ordinario, con la excepción de que el jugador no es penalizado.

Si después de jugarse 30 cartas no hay todavía ningún Profeta, los jugadores son expulsados de la ronda tan pronto cometan un error, esto es, jueguen una carta mala o se declaren sin juego equivocadamente. El jugador expulsado recibe las cartas de penalización correspondientes a su última jugada, que conserva para determinar su puntuación, y abandona la ronda.

Si hay Profeta, se retrasan las expulsiones hasta que se hayan colocado al menos 20 cartas a partir del marcador del Profeta. Como marcadores se utilizan peones de ajedrez, con lo que resulta evidente cuándo es posible la expulsión. Mientras no haya Profeta, se pone un peón blanco cada décima carta colocada en el tablero. Si hay Profeta, se



Se coloca un peón blanco de ajedrez cada décima carta jugada, y un peón negro cada décima carta posterior al marcador del Profeta

pone un peón negro cada décima carta a contar desde el marcador del Profeta. Al derrocar a un Profeta se retiran el marcador de éste y los peones negros.

Así pues, en una ronda puede entrarse y salirse de fases donde son posibles expulsiones. Por ejemplo, si se han colocado ya 35 cartas en la mesa y todavía no hay Profeta, si Pérez juega incorrectamente será expulsado. A continuación López juega correctamente y se declara Profeta. Suponiendo que ahora Gómez jugara incorrectamente, no sería expulsado, pues todavía no se han colocado 20 cartas tras la marca del Profeta.

Una ronda puede finalizar de dos maneras: (1) cuando un jugador se libra de todas sus cartas, o (2) cuando todos los jugadores (excluido el Profeta, si lo hay) hayan sido expulsados.

La puntuación del juego de Eleusis se realiza del modo siguiente:

1. El máximo número de cartas que todavía conserve cualquier jugador (incluido el Profeta) se llama "cuenta máxima". Cada jugador (Profeta incluido) le resta a la cuenta máxima el número de cartas de su mano. Esta diferencia es su puntuación. Si no tiene ninguna carta recibe además una bonificación de cuatro puntos.

2. El Profeta, si lo hay, también recibe una bonificación. Esta es el número de cartas de la línea principal situadas después de su marcador, más el doble del número de cartas laterales situadas tras él, es decir, un punto por cada carta correcta desde su ascenso a Profeta más dos puntos por cada carta errónea a contar desde ese momento.

3. El repartidor tiene puntuación igual a la máxima de los jugadores. Hay una excepción: Si hay Profeta se cuenta el número de cartas (buenas y malas) que procedan al marcador del Profeta y se duplica este número; si la puntuación máxima es mayor que el número resultante, la puntuación del repartidor es este último número.

Si se dispone de tiempo para una segunda ronda, se elige nuevo repartidor. En principio, el juego terminaría después de que todos los jugadores hubieran pasado por un turno como repartidores, pero para ello es fácil que fuera necesario un día entero. Si se desea terminar el juego antes de que todos hayan sido repartidores, lo que se hace es sumar cada jugador las puntuaciones de las distintas rondas, más 10 puntos si no ha sido nunca repartidor; se opera así para compensar el hecho de que las puntuaciones de los repartidores tiendan a ser más altas que el promedio.

La disposición figurativa de las dos páginas anteriores muestra el final de una ronda con cinco jugadores. Pérez era repartidor. La ronda terminó cuando Gómez se deshizo de todas sus cartas. López era Profeta, y terminó con nueve cartas. Rodríguez fue expulsado al jugar indebidamente el 10 de picas; tenía entonces 14 cartas. Alvarez tenía 17 cartas al terminar la mano.

La cuenta máxima es 17. Así que la puntuación de Alvarez es 17 menos 17, o sea, cero. La de Rodríguez es 17 menos 14, es decir, tres. Gómez recibe 17 menos cero, esto es, 17, más una bonificación de cuatro puntos por haberse descartado completamente, así que su puntuación es 21. López tiene 17 menos 9, esto es, ocho, más la bonificación de Profeta, 34 (12 de la línea principal más 11 cartas de líneas laterales situadas tras su marcador), lo que eleva su total a 42 puntos. Es la máxima puntuación de esta ronda. El doble del número de cartas que preceden al marcador del Profeta es 50. El repartidor, Pérez, recibe 42 puntos, porque el menor de los números 42 y 50 es 42.

Se invita a los lectores a observar la disposición de las cartas de la figura, y vean de inferir cuál ha sido la regla secreta. Se trata de una partida estándar, por lo que las condiciones de la regla se refieren solamente a las cartas de la línea principal. La regla correspondiente a esta partida se dará el mes próximo.

He aquí diversos consejos que Abbott ofrece a los noveles jugadores de Eleusis. Como las cartas llegan a ocupar mucho espacio, puede que lo mejor sea ponerlas en el suelo. No hay, desde luego, inconveniente en utilizar una mesa grande o cartas miniatura en una mesa más pequeña. En caso necesario puede romperse la línea principal, recomenzando más abajo desde la izquierda.

Recuérdese que el juego de Eleusis exige al repartidor elegir una regla secreta ni demasiado fácil ni demasiado difícil para alcanzar una puntuación elevada. Como es natural, la bondad de su elección dependerá de la sutileza con que sepa calibrar la habilidad de los jugadores y de la precisión con que gradúe la dificultad de la regla. Ambos cálculos exigen considerable experiencia. Los principiantes tienden, en general, a subestimar la dificultad de las reglas que inventan.

Así, por ejemplo, la regla utilizada para la primera disposición de las cartas es bastante sencilla. Compáresela con esta otra: "Jugar una carta roja, después una carta negra, después una impar, después una par, y repetir cíclicamente". Esta regla es en apariencia sencilla, pero, en la práctica, el cambio de las variables roja-negra a las par-impar la hace más

difícil de descubrir. Abbott señala que, en general, las reglas con condiciones estrictas, en las que solamente una cuarta parte de las cartas jugadas en la mano resultan aceptables, son más fáciles de descubrir que las que presentan condiciones menos estrictas, en las que más o menos la mitad de las cartas resultan aceptables.

No me entretendré detallando en qué forma este juego proporciona modelos de búsqueda de verdades científicas, matemáticas o metafísicas, pues ya lo hice en mi anterior artículo sobre el tema. Tan sólo añadir una fantasía. Quién sabe si Dios, o la Naturaleza, no estarán jugando millares, o quizás un número incontable, de partidas simultáneas de Eleusis con las inteligencias de planetas del Universo entero, maximizando su placer eligiendo en cada caso sistemas de reglas que las mentes inferiores no encontrarán demasiado fáciles ni tampoco imposibles de descubrir si se les da tiempo suficiente. Hay infinitas cartas en reserva; y cuando un jugador es expulsado, siempre hay otros dispuestos a tomar su puesto.

Hay un ir y venir de Profetas y Falsos Profetas, ¿y quién sabría decir cuándo termina una ronda y cuándo comienza otra? La búsqueda de una verdad es un juego apasionante. Y vale la pena tener presente que el juego sería imposible si no estuviesen ocultas sus reglas.

Los lectores que deseen un sistema más completo de instrucciones para jugar a Eleusis pueden enviar 1 dólar a Robert Abbott, quien se las enviará a portes pagados; además serán incluidos en su lista de corresponsales, informándoseles de los desarrollos que el juego experimente en el futuro. La dirección de Abbott es Box 1175, General Post Office, New York. N.Y. 10001.

Se preguntaba, el mes pasado, si la sombra rectilínea de la fotografía del Planetarium McDonnell de St. Louis correspondía a una de las líneas rectas que reglan la superficie de la estructura, que tiene forma de hiperboloide de revolución de una hoja. La respuesta es afirmativa. En aquellos momentos en que la recta que contiene al borde de la sombra interseca al sol, el borde de la sombra en el lateral del planetario coincide con una recta generatriz de la superficie. Esa recta es proyección de un único punto de la circunferencia del techo circular del planetario, y no es proyección de la circunferencia completa. Cuando el sol tenga cualquier otra inclinación el borde de la sombra será una línea curvada, y no será generatriz de la superficie.

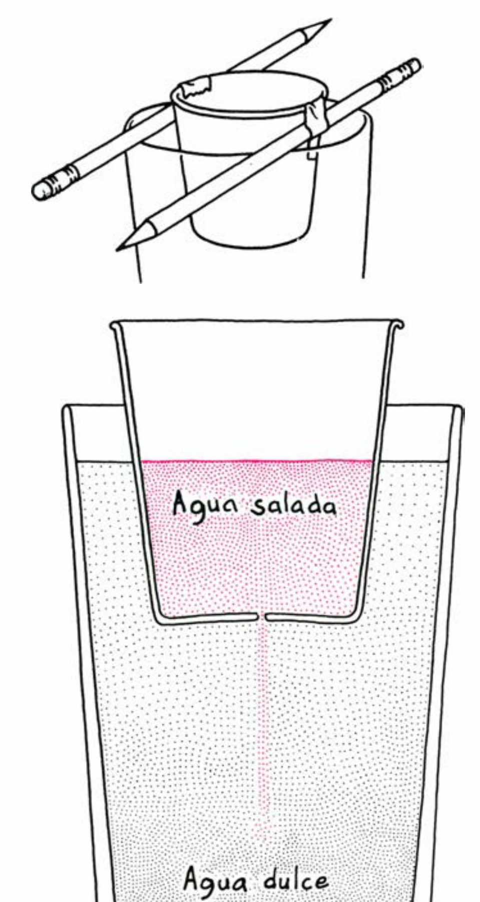
Taller y laboratorio

El manantial salino y otras curiosidades basadas en las diferencias de densidad de los líquidos

Jearl Walker

El hecho de que el agua salada sea más densa que el agua dulce es la base de numerosos experimentos, algunos de los cuales resultan bastante sorprendentes. Por ejemplo, se puede pensar que si se echa una capa de agua salada encima de otra capa de agua dulce, la salada se mezclaría simplemente con la dulce. Sin embargo, si el agua salada se echa de una determinada manera, el resultado no es tan simple; el sistema comienza a oscilar y puede continuar oscilando durante un tiempo bastante largo.

Eche agua fría del grifo en un recipiente alto de cristal, hasta que esté casi lleno. Tome un vaso de papel o de plástico y hágale un agujero en el fondo con un alfiler o algo que se le parezca. Prepare, con agua del grifo y una cucharadita de sal, una solución salina en cantidad aproximadamente igual a la mitad del volumen del vaso de plástico, teniendo cuidado de que la sal se disuelva totalmente en el agua. Añada a dicha solución un poco de colorante doméstico, de manera que después, cuando el experimento esté en marcha, se pueda distinguir fácilmente la solución salina del agua dulce.



Oscilador de sal elemental

A continuación, comience a sumergir el vasito dentro del recipiente alto, vertiendo simultáneamente la solución salina dentro del vasito. Sumerja éste hasta una profundidad en la cual el agua le llegue, por fuera, hasta la mitad y el agua salada del interior alcance aproximadamente el mismo nivel. En este momento se deberá sujetar el vaso, ya que su posición es inestable. Yo lo sujeté con un sujeta-recipientes acoplado a un vástago vertical apropiado, del tipo de los que se usan en los laboratorios de química, pero se puede tratar de mantener quieto el vaso pegándolo con cinta adhesiva a las hojas de dos cuchillos de cocina, que se colocan atravesados en la parte superior del vaso.

Se podría suponer que, en esta situación, la solución salina teñida saldría a través del agujero para mezclarse con el agua dulce, hasta que el nivel de aquélla dentro del vaso fuera tal que la presión hidrostática en el agujero debida a dicha altura fuera la misma desde ambos lados, es decir, desde el lado del agua dulce y desde el lado de la salada; en ese instante, el flujo de solución salina se iría haciendo cada vez más pequeño hasta desaparecer. Eso es lo que puede observarse realmente durante varios minutos, pero el experimento no acaba ahí. Muy pronto, el flujo de agua salada comenzará a hacerse visible de nuevo. A partir de ese momento, y durante un período de unos 15 a 20 segundos, el flujo se reanudará bruscamente, disminuirá y desaparecerá, para volver a empezar de nuevo intensamente.

Lo que ocurre es que, durante el período de tiempo en que el flujo de la solución teñida no resulta visible, el agua dulce está ascendiendo hacia el vaso menor a través del agujero. El sistema se halla en oscilación. Este efecto fue descubierto, en 1970, por Seelye Martin, de la Universidad de Washington.

El período de oscilación depende principalmente del tamaño del agujero y, en menor proporción, de la concentración de sal en el vaso. Mi oscilador continuó toda la noche, con aproximadamente el mismo ciclo, a pesar de la continua reducción de la concentración salina debida a la dilución por mezcla. El oscilador de Martin, construido de hojalata en vez de plástico, duró cuatro días. También repitió el experimento con una jeringa hipodérmica en substitución del vaso; obtuvo así un período de cuatro segundos aproximadamente con una duración del oscilador de 20 ciclos.

La primera parte del flujo se entiende fácilmente. En un principio, el agua salada tiene dentro del vaso la misma altura que el agua dulce del exterior. Dado que la solución salina era más densa que el agua dulce, la presión en el agujero era mayor por la parte de aquélla que por el lado de ésta. La solución salina, por tanto, se vaciará a través del agujero hasta que las presiones se nivelen. La cuestión a resolver es por qué no se estabilizó la solución salina cuando el flujo descendente fue debilitándose hasta detenerse.

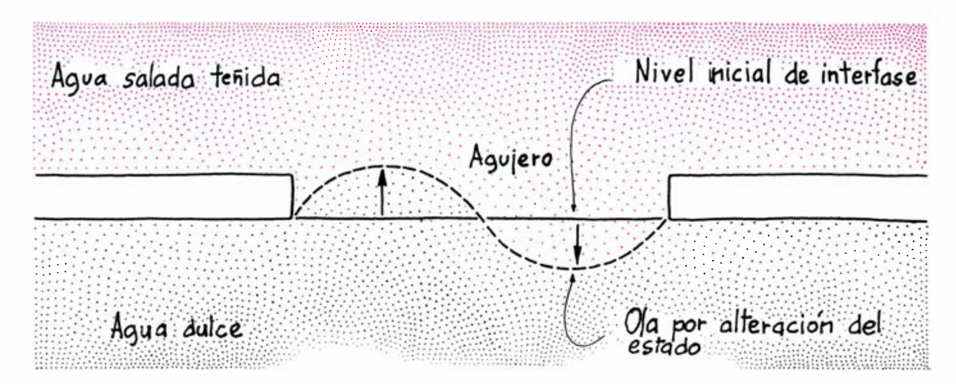
A pesar de que las presiones se hallen igualadas en el agujero, una capa de líquido más denso que esté cubriendo a otra capa de líquido más ligero se encuentra en un equilibrio inestable y, por tanto, expuesta a que cualquier trastorno fortuito, por pequeño que sea, descomponga el sistema. Cualquier alteración de esa índole creará una pequeña ola en la interfase entre las dos clases de líquidos. A causa de la diferencia de densidad, la ola crecerá en amplitud exponencialmente con el tiempo, por lo menos al principio. En consecuencia, porciones del líquido menos denso sobresaldrán cruzando la interfase anterior y porciones del líquido más denso se hundirán. Tal inestabilidad debida a una causa fortuita y la intromisión resultante de cada líquido en la región del otro son las responsables de las oscilaciones del oscilador salino.

Las pequeñas porciones de agua dulce que se introducen en la interfase cruzan rápidamente por el agujero, porque son menos densas que la solución salina que se encuentra a su mismo nivel y al otro lado del agujero. Al parecer, la corriente de agua dulce impide pronto el

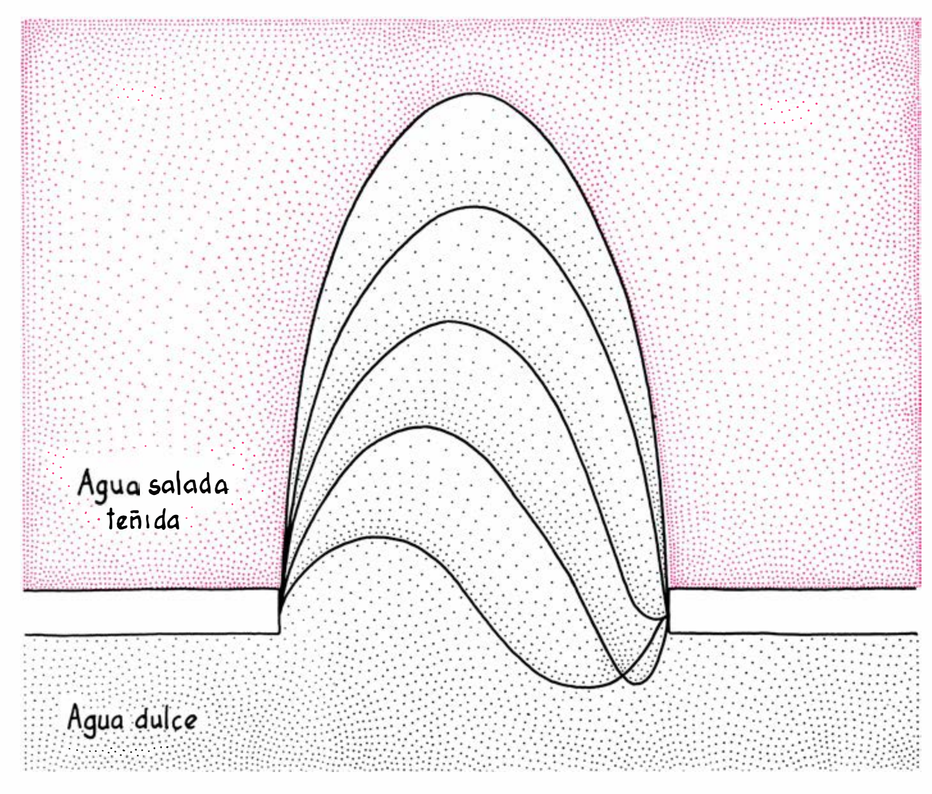
paso de la corriente de agua salada; a partir de entonces, sólo habrá una corriente a través del agujero, que será la corriente de agua dulce en dirección ascendente. Esta incorporación de agua dulce en el vaso mayor irá incrementando gradualmente la altura de líquido en el mismo y, como consecuencia, la presión hidrostática a nivel del agujero. La pérdida de agua en el vaso grande apenas disminuirá el nivel de agua en él, debido a que es mucho más ancho que el vasito de plástico. Puede ocurrir que la presión por el lado de la solución salina llegara a ser lo suficientemente grande a nivel del agujero como para ir dificultando, e impedir por último, la corriente ascendente de agua dulce. En este momento estamos volviendo al punto de partida de un nuevo ciclo. Ahora hay demasiada agua en el vaso pequeño; aparece entonces una corriente de agua salada en dirección descendente durante un rato. Gradualmente, el flujo de dicha corriente va disminuyendo a medida que la presión va haciéndose igual a la altura del agujero. En ese instante, cualquier trastorno fortuito dará lugar a la aparición de una ola en la interfase, que otra vez mandará hacia arriba una corriente de agua dulce. Con una frecuencia que dependerá del diámetro del agujero, del diámetro del vasito y de la viscosidad del agua, la corriente irá alternativamente hacia arriba y hacia abajo. Ya tiene el lector un oscilador salino.

La inestabilidad de la interfase entre un líquido más denso que está sobrenadando a otro líquido menos denso, es decir, cuando la interfase está en equilibrio hidrostático se llama "inestabilidad Rayleigh" (a veces "inestabilidad Rayleigh-Taylor). El oscilador salino es un ejemplo de un sistema que oscila estimulándose a sí mismo, en este caso a través de la "inestabilidad Rayleigh" y el consecuente desarrollo rápido de un trastorno que se produce en la interfase entre los dos líquidos.

Al parecer, los trabajos de Martin es lo único que se ha publicado sobre los osciladores salinos. Tal vez al lector le interese seguir trabajando sobre ellos. Por ejemplo, ¿cómo varía el período de oscilación con respecto al diámetro del agujero? La forma de trabajar de Martin, usando una jeringa hipodérmica con su correspondiente aguja, es muy ventajosa; en efecto, puede contar con agujas de diferentes diámetros fácilmente intercambiables, se conoce exactamente el diámetro de las mismas y la sección circular es mucho más precisa que la de nuestro agujero hecho con un alfiler. El lector puede comprobar la dependencia general del período respecto del diámetro del orificio, entreteniéndose en hacer una serie de agujeros con el alfiler.



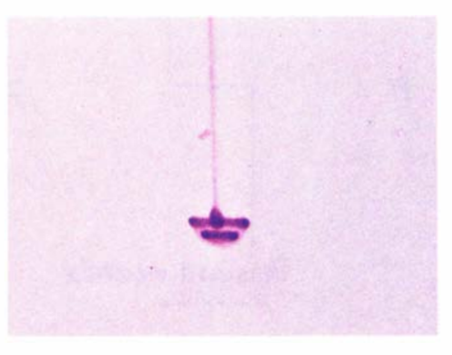
Empieza una oscilación

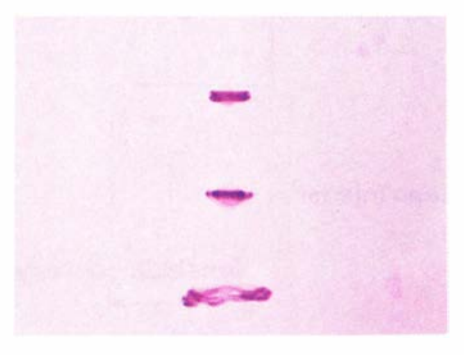


Un chorro ascendente de agua dulce empuja al chorro descendente de agua salada

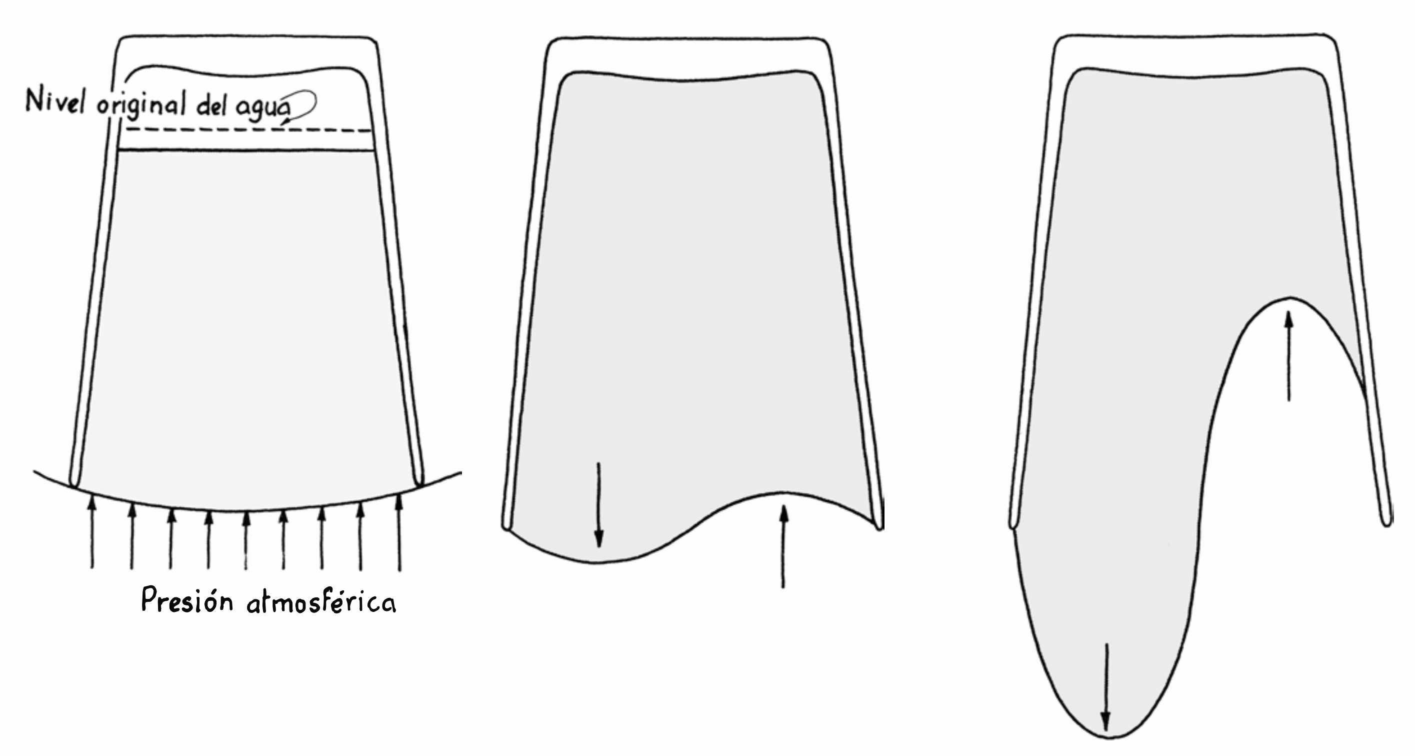
Puede estudiarse la dependencia del período de oscilación con respecto a la concentración de la solución salina. Se parte de una solución salina débil en el vasito y se le va añadiendo sal, teniendo cuidado de que ésta no obstruya el agujero.

También puede el lector ensayar con otros líquidos. No tienen por qué ser solubles uno en el otro, todo lo que se les exige es que uno sea más denso que el otro, y que el más denso se coloque den-





Formas de la corriente descendente



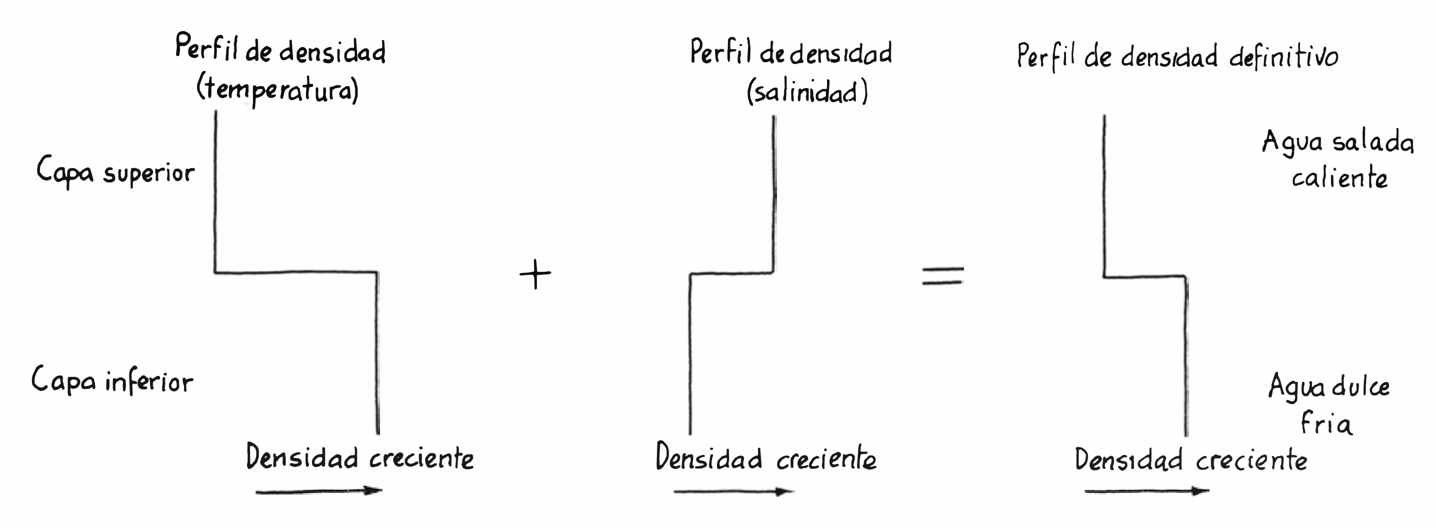
El juego del vaso invertido

La acción conjunta del aire y del agua hacen que ésta caiga

tro del recipiente menor. Yo ensayé con agua y almíbar ligero de cereales (almíbar claro de los que se usan en cocina y repostería); teñí el almíbar de rojo claro antes de verterlo en el vasito. Con un agujero de tamaño adecuado, las oscilaciones del líquido rojo viscoso y del agua dulce fueron casi espectaculares. La estela ascendente se dejaba ver mejor en el oscilador de almíbar que en el oscilador salino. Teñí ligeramente de azul el agua dulce y pude ver con facilidad la estela ascendente de agua dulce que penetraba periódicamente a través del agujero, abriéndose camino hacia arriba por entre el almíbar para luego extenderse por encima de éste.

Cuando empieza la corriente descendente durante el ciclo del oscilador salino, la solución salina teñida parece como si estallase hacia abajo, tomando la forma de un paraguas invertido. Durante la parte del ciclo en la que el flujo es más estable y la corriente fluye sin oscilaciones, el chorro desciende varios centímetros, para separarse luego en anillos concéntricos. Los anillos se parecen a las volutas que hacía mi abuelo con el humo de su puro en los tranquilos días de Aledo, Texas, cuando aquellos anillos de humo eran nuestra única diversión. Los anillos corren uno tras otro, se anudan entre sí y después desaparecen. Este espectáculo se contempla mejor con los anillos de agua teñida, si se usa un recipiente circular de cristal, una taza grande de té, ya que la superficie curva del recipiente aumenta la imagen.

Otro ejemplo de inestabilidad tipo Rayleigh puede observarse en un juego de salón muy común. Se llena parcialmente un vaso de agua, se coloca un trozo de papel sobre el borde y, manteniendo con la mano el papel en su sitio, se invierte el vaso lleno de agua. Después se retira la mano que sostiene el papel; el agua permanecerá en el vaso gracias a dos efectos: primero, a que la tensión superficial entre el agua y el papel y entre el agua y los bordes del vaso ayudan a mantener el papel en su sitio; segundo, porque al descender un poco la columna de agua en los vasos grandes se reduce la presión del aire que permanece atrapado en el vaso. La diferencia de presión del aire entre la parte superior y el fondo de la columna



Perfiles de densidad cuando una capa de agua caliente salada se extiende por encima de otra de agua dulce fría

de agua ayuda también a mantener el agua en su sitio.

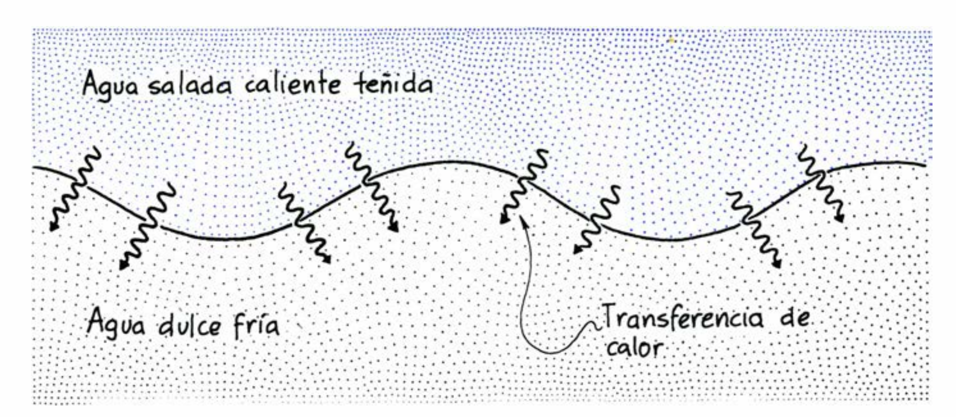
Supongamos que el papel desapareciera de repente sin alterar la superficie del agua, ¿qué ocurriría? El sentido común nos dice que el agua caería; ahora bien, si no hubiera nada que turbara la superficie del agua, no habría ninguna razón para que el agua cayera. Bastaría la diferencia de presión entre la parte alta y el fondo de la columna de agua para mantener el agua en el vaso.

Sin embargo, el agua cae debido a que no permanece en reposo absoluto. Inmediatamente después de que el papel se hubiere retirado, se provocaría una pequeña ola en la superficie del fondo provocada por cualquier accidente fortuito que rompiera el estado de reposo del agua. La ola crecería en amplitud, al principio de forma exponencial con el tiempo, exactamente igual a como habíamos visto en la interfase del oscilador salino, puesto que nos encontramos otra vez ante otro caso de equilibrio inestable tipo Rayleigh. Una burbuja de aire subiría por un lado de la columna de agua, a medida que parte del agua comenzara a bajar por el otro lado, con lo que rápidamente nos encontraríamos con el suelo mojado.

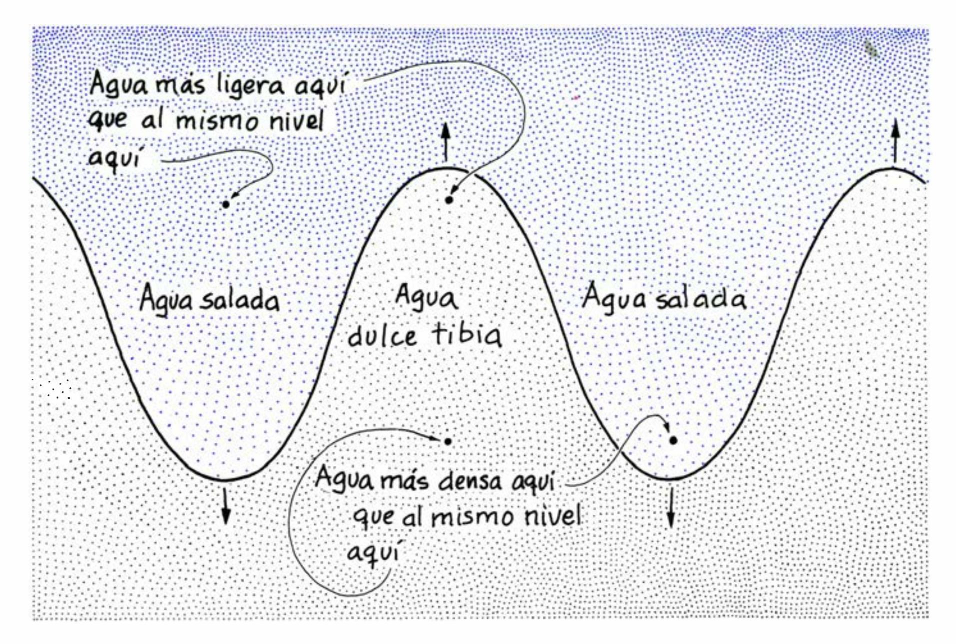
Los "dedos de sal" suponen la existencia de una capa de agua salada sobre otra de agua dulce, similares a las capas existentes en un oscilador salino. Aunque los "dedos" no tienen la fascinación de las oscilaciones, son muy importantes en la microestructura del océano. Consideremos una capa de agua caliente salada sobre otra capa de agua dulce más fría. La sal hace que la capa superior sea más densa que la inferior, pero la diferencia de temperatura compensa de sobra dicha diferencia dando lugar a una situación en la que la capa superior es menos densa que la capa inferior.

Podría suponerse que si una leve alteración enviara parte de la capa superior hacia abajo y parte de la capa inferior hacia arriba, la diferencia de densidades en el conjunto restablecería las interfases originales. Por ejemplo, si la solución salina caliente se moviera hacia abajo sumergiéndose en la capa de agua dulce fría, se encontraría rodeada de un líquido más denso que ella y flotaría de nuevo en dirección ascendente.

Pero si, debido a un trastorno en la interfase, se produce la mezcla de líquidos antes referida, rápidamente se lleva a cabo un intercambio de calor entre la solución caliente y el agua fría. La solución caliente desprende calor al moverse hacia abajo y el agua fría lo absorbe en su desplazamiento hacia arriba. A consecuencia de lo cual, la parte de solución salina que se introduce en la capa inferior se enfría de repente, haciéndose más densa que el agua que la rodea y, en vez de volver a flotar hacia arriba, sufre un impulso hacia abajo. De igual manera, la parte de agua fría que asciende introduciéndose en la capa superior se calienta rápidamente haciéndose menos densa que la solución salina caliente que la rodea y acelerando su movimiento hacia arriba. El trastorno inicial registrado en la interfase entre las dos capas queda, por tanto, agrandado produciendo "dedos" de solución salina que se extienden hacia abajo y agua dulce que empuja hacia arriba. Con el tiempo, los "dedos" se irán alargando por sí solos hasta un punto en donde se doblen para formar una capa de mezcla de agua con una salinidad y temperatura que serán intermedias entre las de las capas superior e inferior.



Comienzo de un desequilibrio fortuito en la interfase de las capas



Incremento de la ola a medida que el calor se difunde más rápidamente que la sal

Los "dedos de sal" no se observan directamente en el océano, aunque sí han sido detectados los perfiles resultantes de temperatura y salinidad. Por ejemplo, las aguas salinas y calientes del Mediterráneo reúnen las condiciones necesarias para formar "dedos de sal" a su paso por el Estrecho de Gibraltar, al fluir sobre la corriente de agua menos salada y más fría procedente del Atlántico. Tomas de muestras hechas para medir la salinidad y la temperatura en función de la profundidad revelan unos perfiles en forma de escalones en los que la salinidad y la temperatura decrecen con la profundidad. Dichos escalones tienen cada uno una capa cuya alta salinidad y temperatura son más elevadas que la capa inferior, y existe entre ambas una capa de valores intermedios.

El lector puede producir "dedos de sal" muy fácilmente en su cocina vertiendo cuidadosamente agua caliente salada (teñida con colorante doméstico) sobre agua dulce más fría. Al echar el agua, se debe tratar de reducir al mínimo la turbulencia que se origina en la mezcla de los dos tipos de agua. Si el recipiente es un vaso común, yo lo in-

clino para que el agua caiga a sólo un centímetro o menos. Si el recipiente es mayor, bajo el vaso del agua salada hasta casi el nivel del agua dulce. También se puede hacer flotar un trozo de madera en la superficie del agua dulce y echar la salada sobre el trozo de madera.

Al principio, la interfase entre el agua caliente salada y teñida y el agua dulce limpia sólo indica algo de mezcla. En pocos minutos se desarrollarán en la interfase los "dedos de sal", de uno a cinco centímetros de largo y de un milímetro de ancho, más o menos. Su duración puede variar desde varios minutos hasta unas horas. Para observar mejor los "dedos" hay que enfocarlos con una linterna a través del agua y en dirección a los ojos.

La clave de los "dedos de sal" es la diferencia en la velocidad de difusión del calor y la sal. El calor que hizo la capa superior más ligera, a pesar de su concentración salina, se mueve hacia el agua dulce más fría unas 100 veces más deprisa que la sal. La rápida transmisión del calor permite que el agua salada de movimiento descendente se torne de repente más pesada que su entorno, y que el agua dulce de movimiento ascendente resulte más ligera que el agua que la rodea.

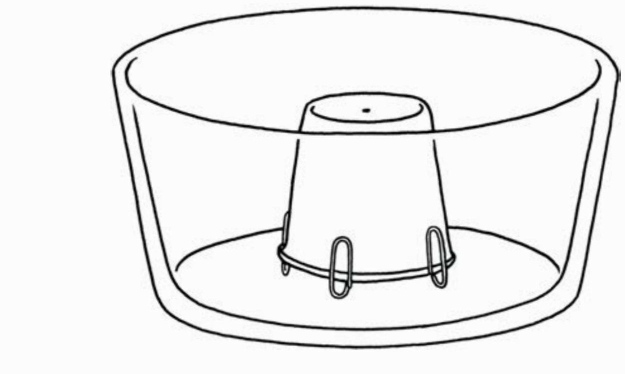
Se puede hacer también el mismo tipo de "dedos" convectivos con soluciones de azúcar y sal. Para que la solución de azúcar sea menos densa que la solución de sal, se echa una cantidad de azúcar inferior a la de sal en un volumen determinado de agua. Tiña la solución azucarada, de manera que pueda seguirse su movimiento; vierta suavemente la solución de azúcar sobre la de sal en un recipiente. Puesto que las dos soluciones están a la temperatura del agua fría del grifo, estos "dedos de sal" son más fáciles de hacer en la cocina que el experimento anterior, en el que el enfriamiento rápido del agua caliente puede constituir un problema.

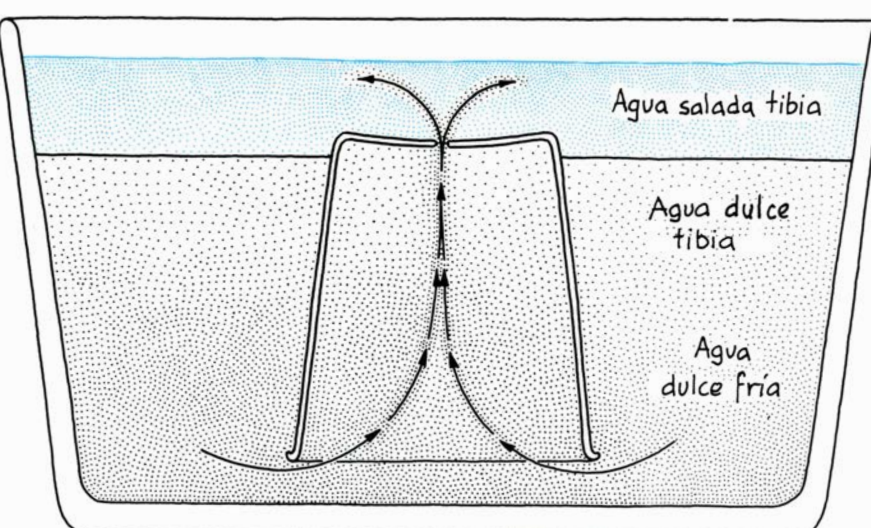
Una vez más, como sucedía con la disposición del agua caliente salada y el agua corriente fría, las capas de solución de azúcar y sal deberían ser estables, porque la capa superior es la menos densa. Sin embargo, cualquier movimiento inicial se verá incrementado en virtud de la diferencia que existe en la velocidad de difusión: la sal se difunde más rápidamente que el azúcar. Por ello, cuando el movimiento inicial envía una pequeña cantidad de solución de azúcar hacia abajo, la sal de los alrededores ocupa su lugar antes de que el azúcar pueda hacerlo. La porción de agua con más sal es de repente más densa que su entorno, y precipita hacia abajo produciendo un "dedo". De forma similar, una pequeña porción de agua salada de movimiento ascendente pierde sal más deprisa que gana azúcar; por ello, resulta más ligera que su entorno y sufre una aceleración hacia arriba para producir un "dedo".

La acción conjunta de sal y calor en la densidad del agua de mar y la diferencia en su velocidad de difusión conducen también a una curiosa fuente "perpetua" que el lector puede simular en su cocina. En un océano tropical, el agua próxima a la superficie puede estar relativamente caliente y salada, mientras que el agua próxima al fondo está relativamente fría y sin sal. Imagínese una tubería colocada verticalmente casi hasta el fondo y una bomba que se usaría, inicialmente, para subir agua del fondo a la superficie. En teoría, la bomba podría quitarse y el manantial continuaría por sí solo. Dado que una tubería difícilmente podría alterar las características del océano, este chorro continuaría fluyendo "siempre".

Para entender cómo se autobombea el manantial, consideremos una pequeña cantidad de agua que empieza a subir por la tubería. A medida que esa pequeña cantidad de agua fría ascienda, irá ganando calor gradualmente del agua más caliente que rodea a la tubería a la misma profundidad. La porción de agua interior sería entonces más ligera que el agua más salada a la misma profundidad y temperatura, pero exterior a la tubería. En consecuencia, el agua de la tubería aceleraría su ascenso. Teniendo en cuenta que la pequeña porción de agua se iría calentando continuamente de esta forma a lo largo de su ascenso, siempre resultaría más ligera que el agua del exterior y seguiría acelerándose en su ascenso. En principio, el agua subiría un par de metros por encima de la superficie del océano formando una fuente perpetua.

Hace algunos años, Henry M. Stommel, del Instituto de Tecnología de Massachusetts, intentó poner en marcha la fuente mencionada en las profundidades oceánicas inmediatas a la isla de la Martinica. Con unos 1000 metros de manguera flexible, de 18 centímetros de diámetro, él y sus colegas consiguieron hacer una fuente de unos 60 centímetros de altura, pero dudaban de que fuera el resultado de diferencias de densidad. La punta de la manguera estaba sujeta a un flotador que subía y bajaba con las olas, estirando la parte alta de la manguera y creando una especie de acción de bombeo. Stommel piensa que la fuente se debe a la acción de bombeo.





Disposición para generar una fuente salina

Para simular la fuente incesante en su cocina, llene parcialmente un recipiente ancho con agua del grifo. Haga un agujero en el fondo de un vaso de papel o de plástico y póngalo en el agua boca abajo. Después, eche cuidadosamente agua caliente sobre el agua fría en el exterior del vaso. Trate de evitar la mezcla excesiva de las dos capas. Continúe vertiendo agua hasta que la adición de agua caliente haya forzado al agua fría a subir dentro del vaso invertido hasta el nivel del agujero.

Vierta entonces una capa fina de agua caliente salada sobre la superficie del agua caliente y también sobre el vaso, teniendo cuidado otra vez de evitar la mezcla excesiva. Ponga una o dos gotas de colorante justo sobre el agujero del vaso invertido. El colorante le mostrará que una fuente lenta de agua fluye desde el agujero a través de la capa alta de agua; se obtiene, pues, una réplica de la fuente de sal oceánica. El agua que sube a través del vaso se va calentando hasta que alcanza aproximadamente la misma temperatura que el agua situada a la misma profundidad fuera del vaso, pero no tiene la sal que posee el agua de fuera, y es, por tanto, más ligera que ella. Ello explica que el agua del vaso se sienta obligada a ascender a través del agujero. La fuente continuará manando hasta que el calor y la sal se distribuyan más regularmente.

Como nota final de los efectos debidos a la diferencia de densidades del agua salada y la dulce, permítanme describir el empuje que recibe un barco cuando las compuertas de la última esclusa del Canal de Panamá se abren. A medida que un barco se va acercando a la orilla del Pacífico del Canal, va descendiendo progresivamente, a través de las esclusas, hasta el nivel del mar. La mayor parte del agua de estas esclusas es agua dulce aportada por los lagos de Panamá, que a su vez se nutren del agua de lluvia.

En cada esclusa el barco tiene que esperar hasta que el nivel del agua desciende hasta el de la siguiente esclusa. Entonces, la presión del agua en los dos lados de la compuerta que separa las dos esclusas es la misma y la compuerta se abre con facilidad.

En la última esclusa, el agua del otro lado es la salada del océano. Dado que el agua salada es más densa que el agua dulce, la presión de equilibrio en la puerta se alcanza cuando el agua dulce está todavía un poco por encima del nivel del océano. Cuando la compuerta se abre, esta altura de más del agua dulce corre hacia el mar arrastrando el barco con ella en un breve y gratuito paseo.

Libros

Introducción a la conducta animal y dos valiosos testimonios de Konrad Lorenz y Karl v. Frisch

Luis Alonso

INTRODUCCIÓN A LA CONDUCTA ANIMAL, por Aubrey Manning, trad. cast. de Natividad Sánchez Sáinz-Trápaga, Alianza Editorial, Madrid, 1977. Nuestra historia cultural tan densa en cuanto a tratadistas de la ética, del *éthos* o conducta y talante humanos, con su disciplina antaño ancilar, y hoy plenamente emancipada, la psicología, desconoce por completo la existencia de otro *éthos,* más primitivo y fundante, el *éthos animal* y la ciencia que lo considera, la etología o estudio de la conducta animal. No vale argumentar que es una ciencia joven. Más joven era años atrás en que un aguerrido profesor acometió el empeño de enseñarla, hasta que hubo de desistir porque las arcas de la Universidad no podían permitirse semejante dispendio.

En sentido amplio, en la definición de etología entra cualquier actividad del organismo. La conducta, acotando ya el alcance estricto del término etología, comprende todos aquellos procesos a través de los cuales un animal percibe el mundo externo y el estado interno de su cuerpo, y responde a las variaciones que registra. Una definición plenamente skinneriana, se diría. Pero en biología, más que de la relación estímulo-respuesta, hay que hablar del encabalgamiento siguiente: estructura-función-medio ambiente. La etología recaba de la fisiología (de la neurofisiología, principalmente) los elementos constitutivos de sus sistemas y aparatos y debe a la ecología la ubicación en tiempo y lugar del funcionamiento del organismo. Escribe Margalef a este último respecto: "El comportamiento de especies que conviven aparece integrado de acuerdo con la naturaleza de las interacciones entre las especies en cuestión, sean depredador y presa, o sean especies que compiten por unos mismos recursos y que encuentran la forma de convivir, complementando de algún modo sus actividades" (Ecología, 1974).

El "anecdotario" animal es antiquísimo. La literatura griega tiene abundantes muestras ya. Y junto a las fábulas y otras referencias (Aristófanes, por ejemplo) encontramos una maraña sutil de literatura esotérica en la que se magnifican distintos comportamientos animales y se interpretan la acción y el futuro humanos de acuerdo con el vuelo de las aves, su canto, etcétera. De una manera voluntariamente científica, la llamada "historia natural" se ocupaba también de la actividad zoológica con detalles bastante finos (Buffon, Audubon, Brehm Fabre), conformándose la disciplina que hoy reconocemos por etología merced a los trabajos más recientes de Jacques Loeb, Herbert S. Jennings, Tinbergen, K. Lorenz y Von Frisch.

Aubrey Manning agrega al final de su obra una copiosa bibliografía en la que se recogen prácticamente casi todos los trabajos, aunque se haya dejado algunos importantes en el fichero (August Krogh y H. Kalmus, ambos especialistas en etología de artrópodos), que han sentado las bases del estudio del comportamiento animal.

Los animales se mueven en todo momento en instinto de supervivencia. Para ello deben resolver los mismos problemas fundamentales; así, periódicamente, deben abastecer a la máquina metabólica (comer y beber), deben luchar contra afanes predatorios de otras especies o de otros individuos de su misma especie, deben mantenerse limpios y sanos y deben asistir a las necesidades de mantenimiento de la especie (reproducción). Contra lo que pudiera parecer, no existe en ello uniformidad absoluta, en el sentido de que siempre se hizo así. Porque no siempre se hizo así. El comportamiento del animal se ha ido modificando de acuerdo con su evolución, y los mismos factores que inciden en ésta (migración, mutación y selección), inciden en la etología. Los condicionantes genéticos o hereditarios dejan, pues, sentir su peso, junto con los ecológicos y los fisiológicos.

En la conducta del animal encontramos unos patrones de comportamiento fijos, verbigracia, los "hábitos" de ingeniería animal (la construcción del nido). Vienen luego los llamados estímulos clave o cruciales. El animal reacciona ante relativamente pocos estímulos del atrayente panorama que le ofrece el medio; además, no siempre reaccionará igual ni ante los mismos estímulos (depende de qué condiciones priven en su estado: agresividad, reproducción, etcétera). A este propósito resulta esclarecedor el capítulo cuarto, dedicado a la motivación. Motivación es impulso o tendencia dominante de respuesta ante un determinado estímulo. Comporta subsidiariamente una serie de cambios fisiológicos relacionados con la emoción: resecamiento de la boca, sudoración, aceleración del ritmo cardíaco, y otros. La emoción se mide a posteriori por la conducta observada en el animal: "si un animal ataca a otro atribuimos esto a una motivación agresiva; si come, atribuimos esto a una motivación alimentaria, y así sucesivamente". Otra situación importante en este ámbito es la de los "estímulos supranormales"; nos referimos a los estímulos que provocan una respuesta más intensa que el objeto estímulo natural. En prueba de la existencia de éstos suele citarse el trabajo de Magnus sobre la mariposa fritilaria plateada: los machos son atraídos por las hembras en virtud del relampagueo naranja que producen sus alas al volar; Magnus consiguió atraer a los machos haciendo girar un tambor que emitía una imagen parecida a las alas a distintas velocidades: la frecuencia de aleteo de una hembra fritilaria es aproximadamente de 8 batidas por segundo, pero los machos mostraban respuestas más intensas cuanto más de prisa batían las alas del modelo, cuya frecuencia llegó hasta 75 por segundo.

Subyacente a este comportamiento fisiológico a nivel de sistemas y órganos se encuentra el comportamiento bioquímico determinante del anterior. En particular, las hormonas. Los cambios fisiológicos registrados en un animal proceden directa o indirectamente de los cambios nerviosos y endocrinos, y de sus interacciones. Las hormonas sexuales o gonadotropinas ejercen una incidencia obvia en el comportamiento. Cástrese a un vertebrado en edad juvenil y modificará radicalmente su conducta. O auméntese o disminúyase la concentración habitual de una hormona en el torrente sanguíneo y las manifestaciones exteriores (verbigracia, cortejo, plumaje, etcétera) cambiarán a ojos vista. A veces con efecto contraproducente.

Cierran la obra tres capítulos (el octavo, el noveno y el décimo) dedicados al aprendizaje y a la organización social, aspectos muy vistosos por la serie de comparaciones inmediatas que surgen con la especie y la sociedad humanas

(sobre todo, si los patrones se fijan de acuerdo con niveles superiores de evolución: los primates). El lector de la revista recordará a este propósito dos artículos publicados recientemente: el de Egozcue, sobre las afinidades genéticas de los primates y el de Gray Eaton, sobre la organización social de los macacos japoneses. Ambos compendian los adelantos registrados después de la publicación en su original inglés de la obra de Aubrey Manning.

Pero los progresos de la etología no acaban ahí. Algo dejó entrever Tinbergen cuando avanzó su teoría de las "actividades de desplazamiento", que consisten en patrones de conducta cuya característica común más notable es su aparente irrelevancia para la situación en que aparecen. Es decir, no viene a cuento que el animal obre así, porque no existe el estímulo externo correspondiente. El animal se encuentra entre dos tendencias opuestas o bien se halla en una situación de frustración; por ejemplo, la paloma sedienta que se ve imposibilitada de llegar hasta el bebedero porque se ha interpuesto un cristal a modo de obstáculo; en respuesta, ella se dedicará a picotear el suelo.

La teoría de la actividad del desplazamiento puede asumirse en un modelo de interpretación de la conducta animal mucho más general y comprehensivo. Se puede enfocar desde la llamada "teoría de la decisión". No se trata sólo de analizar una situación límite, propiamente dicha. Sino de examinar el comportamiento desde una pauta digamos "inteligente". En un momento dado, el animal tiene ante sí un panel de posibles actividades que acometer: alimentarse, cortejar, dormir y otros. No puede hacerlos todos a la vez porque los movimientos requeridos suelen ser mutuamente antagónicos. Por consiguiente, tiene que existir algún proceso o mecanismo que dicte qué actividad goza de prioridad en ese instante dado. A dicho proceso se conviene en darle el nombre de proceso de decisión.

Las decisiones tomadas por los animales pueden ser decisiones sobre continuar o no la actividad que tienen entre manos o comenzar otra. Hay que tomar decisiones también sobre posibilidades alternativas. En cada caso debe "considerar" toda una serie de factores. Valga el ejemplo de la gaviota *(Larus argentatus)*, que está incubando sus huevos; la hembra permanecerá así hasta que venga el macho a relevarla, pues ambos se encargan de tal protección, porque de lo contrario el nido caería en las garras de los depredadores. El macho se halla en búsqueda de alimento, operación que le

lleva algunas horas; puede ocurrir, sin embargo, que haya sufrido algún percance (desde no encontrar comida y tener que alejarse más, hasta caer en manos de un científico que esté llevando a cabo alguna campaña). ¿Qué hará la hembra? El dilema que se le presenta es: esperar que vuelva su pareja o salir en búsqueda de alimento, necesidad esta última que se va haciendo más apremiante. La situación se complica por cuanto la hembra debe asegurarse el regreso, cuestión que no es nada fácil porque varía en relación con el tiempo, el estado del oleaje y la hora del día, y otros problemas más, derivados, sobre todo, de la tendencia a especializarse en las distintas tareas. Ante ello, el etólogo debe plantearse una doble pregunta: ¿qué decir de los procesos de decisión imbricados en ese comportamiento? y ¿hasta qué punto pueden los animales tomar decisiones de tal complejidad? Las respuestas no son nada sencillas. Están ahí como un reto permanente. Una vía de investigación sería relacionar y sopesar valores del análisis coste-beneficio, es decir, los intereses que arriesga al obrar de una manera en razón de las probabilidades de éxito. ¿Qué éxito, a su vez, es el predominante?, ¿el suyo?, ¿el de su cría?, ¿el de la especie? De la comprensión de estos interrogantes depende un conocimiento más profundo del comportamiento del animal en las circunstancias cambiantes y en la evolución.

HABLABA CON LAS BESTIAS, LOS PECES Y LOS PÁJAROS, por Konrad Lorenz, Editorial Labor, Barcelona, 1975. La etología recibió su espaldarazo académico cuando en 1973 se concedió a su máximo exponente, Konrad Lorenz, el premio Nobel de Medicina y Fisiología. Lo que empezó siendo una diversión infantil, un regalo paterno en las efemérides más significativas, se fue convirtiendo en pasión y simpatía intelectual. Simpatía en el sentido etimológico del término, es decir, un conocimiento objetivo y distante que se enriquece con las aportaciones cercanas de una vida, y casi unos sentimientos, en común. Este libro espléndido es un retazo vivo de una pasión y una observación atenta y diaria. Los animales no pertenecen a un reino especial, a donde haya que desplazarse para estudiarlos de lejos, sino que son los inquilinos habituales de la casa, con sus nombres, con su talante particular. "Cuenta la historia que el rey Salomón podía hablar con las bestias, las aves, los peces y los gusanos. También puedo hacerlo yo."

Lorenz se procuró una formación de base sólida. Se doctoró en Medicina (Viena, 1928) y en Zoología (1933), especializándose en el campo de la anatomía comparada, que le permitía llegar a explicar mejor los distintos patrones de estructura y comportamiento. Al año de fundarse la Sociedad Alemana de Psicología Animal, esto es, en 1937, pasó a desempeñar la dirección adjunta de la revista *Zeitschrift für Tierpsychologie*, que se convertiría en la publicación pionera de los estudios etológicos. Tras la Segunda Guerra, y de nuevo en su Austria natal, dirigió durante dos años el Instituto de Etología Comparada. Al principio centró su atención en la naturaleza de los actos instintivos, en particular, su fundamentación en el sistema nervioso. Con Niko Tinbergen demostró cómo las distintas formas de conducta se armonizan en una sola secuencia de acción. Muchas de sus impresiones han servido de vías de estudio posteriores específicas. Así, un campo abierto por él fue el de la comunicación. Hoy el tema concreto de las ferormonas, que son ciertas substancias secretadas por un animal que condiciona el comportamiento de otros de la misma especie, está experimentando notables progresos (por ejemplo, el rechazo de una serpiente hembra de la especie *Thamnphis sirtalis* recién cubierta por otro macho ante la "señal" de dominio dejada por el primer copulador). O los capítulos de la etología asociados al principio de territorialidad, cuyo fundamento esbozó al estudiar la agresividad. El animal vive "en su territorio", en él se siente seguro y fuerte, y en él guarda la jerarquía necesaria para la supervivencia de la especie. Apartados concretos de ese aspecto general hoy en estudio son, por ejemplo, cómo enjuician los animales la distancia, cuál es su profundidad de visión. Con el tiempo, Lorenz fue echando cada vez más puentes entre el comportamiento animal y el humano. En su obra capital a este respecto, *Sobre la agresión*, avanza ya lo que ha venido siendo estribillo en estos últimos años: la agresión, tanto en los animales como en el hombre, es el resultado de una tendencia hereditaria, espontánea, cuyas propiedades son en gran parte las mismas que las de los impulsos biológicos de comer y beber. Otro gran biólogo, Ramón Margalef, se encarga, en su traducción, de que la obra llegue al lector con el rigor, la ingenuidad y la chispa humana con que salió de las manos de Lorenz.

LA VIDA DE LAS ABEJAS, por Karl von Frisch, Editorial Labor, Barcelona, 1976. A veces, una obra vale toda una vida. No siempre se reconoce de inmediato, porque, a lo mejor, la atención social vigente está prendada de otras ramas del saber. Cuando Frisch escribió por vez primera su *Vida de las abejas* (1927)

Europa asistía absorta al nacimiento revolucionario de la nueva física atómica. Y nuestro autor, con la humildad de los grandes, no hace justamente un propósito de renovación: "Pero, ¿hacía falta la aparición de un nuevo libro relativo a las abejas? Ahí está la maravillosa obra de Maeterlinck, *La vida de las abejas*, o de Bonsels, *Maya la abeja*, ambas plenas de notables observaciones y un verdadero encanto para el lector versado" (Prólogo a la primera edición, 1927). Algo nuevo tuvo que aportar Frisch para que, profundizando en esa investigación, sin apenas correcciones del libro inicial, le concedieran en 1973 el premio Nobel de Medicina y Fisiología por sus estudios sobre el comportamiento social y el lenguaje de las abejas, galardón que compartió con Konrad Lorenz y Niko Tinbergen.

El lector puede preguntarse con buen criterio si da para tanto un tema tan específico. Como otros científicos austriacos de origen judío, Von Frisch se las tuvo que ver con el régimen nacionalsocialista. Y hubiera sido encarcelado cuando menos de no sospechar el Ministerio alemán de abastecimientos que sus investigaciones sobre las abejas eran de gran interés económico, con lo que se pospuso su "dimisión" de la cátedra de Zoología de la Universidad de Munich, en donde estaba enseñando. Esa anécdota de su vida puede indicar hasta qué punto resultan provechosos ciertos trabajos presuntamente inútiles.

Apenas cumplidos los veinte años, comenzó a estudiar la ceguera de las abejas para el color, según se creía hasta entonces. Con un sencillo experimento, que en el libro aparece pormenorizado, demostró que eran ciegas para el rojo, pero no para otros colores. Con tenacidad, Von Frisch fue estudiando los distintos sentidos de estos insectos y de otros animales inferiores, sobre todo de los peces.

Le llamó en seguida la atención la rauda comunicación que existía en los pequeños artrópodos. Si colocaba agua azucarada en un recipiente y lo descubría una abeja, de inmediato acudía toda una expedición constituida por miembros de la misma colmena. Ese morse ininterpretable hasta entonces era la danza, cuyos pasos y movimientos se cambiaban en razón de la distancia del néctar. Como en todo lenguaje, también en el de las abejas existen unas estructuras comunes y unos dialectos que matizan el patrón general. Tales insectos poseen un código convencional, que las distintas subespecies y razas acomodan a "su nacionalidad". Además, según el nivel de desarrollo idiomático así será el nivel, alto o bajo, de la organización social de la colmena.

Bibliografía

Los lectores interesados en una mayor profundización de los temas expuestos pueden consultar los trabajos siguientes:

EL TRATAMIENTO METABOLICO DE LAS ENFERMEDADES DEL CORAZON

ON THE EQUILIBRIUM OF HETEROGENEOUS SUBSTANCES. J. W. Gibbs (1875). Collected Works, vol. 1. Longmans, Green. New York, 1928.

STRESS AND CELLULAR FUNCTION, H. Saborit. J. B. Lippincott. Philadelphia, 1959.

EFFECTS OF THE POLARIZING SOLUTION GLUCOSE-INSULIN-POTASSIUM (G-I-K) ON THE ELECTROCARDIOGRAM ELECTROLITES AND MORPHOLOGY OF ACUTE EXPERIMENTAL MYOCARDIAL INFARCTION, D. Sodi-Pallares, A. Bisteni, J. A. Medrano, A. De Micheli, J. Ferrer, R. Contreras y E. Zetina-Mora, en *Abstracts IV World Congress of Cardiology,* México, 1962.

EXPERIMENTAL CARDIOVASCULAR DISEASES. H. Seyle. Parte I. Springer-Verlag, New York, Heidelberg, Berlin, 1970.

EFFECT OF GLUCOSE-INSULIN-POTASSIUM INFUSION ON MYOCARDIAL INFARCTION FOLLOWING EXPERIMENTAL CORONARY ARTERY OCCLUSION. P. R. Moroko, P. Libley, B. E. Sobel, C. M. Bloor, H. D. Sybers, W. E. Shell, J. W. Covell y E. Braunwald, en *Circulation,* vol. 45, n.º 1160, junio, 1972.

ESTRELLAS DE RAYOS X EN CUMULOS GLOBULARES

X-RAY-EMITTING DOUBLE STARS. Herbert Gursky y Edward P. J. van den Heuvel en *Scientific American,* vol. 232, n.º 3, págs. 24-35; marzo, 1975.

X-RAY BINARIES IN GLOBULAR CLUSTERS. George W. Clark en *Astrophysical Journal (Letters),* vol. 199, págs. L143-L145; 1 de agosto de 1975.

PARTICULAS FUNDAMENTALES CON ENCANTO

QUARKS WITH COLOR AND FLAVOR. Sheldon Lee Glashow en *Scientific American,* vol. 233, n.º 4, págs. 38-50; octubre, 1975.

OBSERVATION IN e^+e^- ANNIHILATION OF A NARROW STATE AT 1865 MEV/c^2 DECAYING TO $K\pi$ AND $K\pi\pi\pi$. G. S. Abrams, M. S. Alam, A. M. Boyarski, M. Breidenback, W. C. Carithers, W. Chinowsky, S. C. Cooper, R. G. DeVoe, J. M. Dorfan, G. J. Feldman, C. E. Friedberg, D. Fryberger, G. Goldhaber, G. Hanson, J. Jaros, A. D. Johnson, J. A. Kadyk, R. R. Larsen, D. Lüke, V. Lüth, H. L. Lynch. R. J. Madaras, C. C. Morehouse, H. K. Nguyen, J. M. Paterson, M. L. Perl, I. Peruzzi, M. Piccolo, F. Pierre, T. P. Pun, P. Rapidis, B. Richter, B. Sadoulet, R. H. Schindler, R. F. Schwitters, J. Siegrist, W. H. Tanenbaum, G. H. Trilling, F. Vannucci, J. S. Whitaker y J. Wiss en *Physical Review Letters,* vol. 37, n.º, 5, págs. 225-259; 2 de agosto de 1976.

RADAR DE IMAGENES LATERALES AEROTRANSPORTADO

AN INTRODUCTION TO SYNTHETIC-APERTURE RADAR. William M. Brown y Leonard J. Porcello en *IEEE Spectrum,* vol. 6, n.º 9, págs. 52-62; septiembre, 1969.

SYNTHETIC INTERFEROMETER RADAR FOR TOPOGRAPHIC MAPPING. L. C. Graham en *Proceedings of the IEEE,* vol. 62, n.º 6, págs. 763-768; junio, 1974.

SYNTHETIC APERTURE RADAR. Dirigido por John Kovaly. Artech House, 1976.

REMOTE SENSING. Allen L. Hammond en *Science,* vol. 196, n.º 4289, págs. 511-516; 29 de abril de 1977.

ESTRUCTURA Y FUNCIONAMIENTO DE LOS ANTIGENOS HISTOCOMPATIBLES

BIOLOGY OF THE MOUSE HISTOCOMPATIBILITY-2 COMPLEX. Jan Klein. Springer-Verlag, 1975.

STRUCTURE AND SIGNIFICANCE OF β_2-MICROGLOBULIN. Bruce A. Cunningham en *Federation Proceedings,* vol. 35, n.º 5, págs. 1171-1176; abril, 1976.

SURFACE MODULATION IN CELL RECOGNITION AND CELL GROWTH. Gerald M. Edelman en *Science,* vol. 192, n.º 4236, págs. 218-226; 16 de abril de 1976.

LA SOLUCION DEL MAPA DE CUATRO COLORES

EVERY PLANAR MAP IS FOUR COLORABLE: PART 1, DISCHARGING. Kenneth Appel y Wolfgang Haken en *Illinois Journal of Mathematics,* vol. 21, n.º 3; septiembre, 1977.

EVERY PLANAR MAP IS FOUR COLORABLE: PART 2, REDUCIBILITY. Kenneth Appel, Wolfgang Haken y John Koch en *Illinois Journal of Mathematics,* vol. 21, n.º 3; septiembre, 1977.

COMO EMPEZO LA EDAD DEL HIERRO

A STEEL TOOL OF THE FOURTH CENTURY B.C. FROM AL MINA IN SYRIA. Tamara S. Wheeler, James D. Muhly y Robert Maddin en *Levant,* vol. VIII, págs. 107-112; 1976.

AN IRON ADZE OF THE FIFTH-FOURTH CENTURIES B.C. FROM AL MINA. James D. Muhly, Tamara S. Wheeler y Robert Maddin en *Levant,* vol. IX, págs. 156-161; 1977.

ALUCINACIONES

MESCAL AND MECHANISMS OF HALLUCINATIONS. Heinrich Klüver. The University of Chicago Press, 1966.

HASHISH AND MENTAL ILLNESS. J. J. Moreau. Raven Press, 1973.

HALLUCINATIONS: BEHAVIOR, EXPERIENCE AND THEORY. Dirigido por Ronald K. Siegel y Louis Jolyon West. John Wiley & Sons, Inc., 1975.

JUEGOS MATEMATICOS

THE LOGIC OF SCIENTIFIC DISCOVERY. Karl R. Popper. Hutchinson of London, 1959. (Trad. castellana de Edic. Tecnos, Madrid.)

THE JUSTIFICATION OF INDUCTION. Dirigido por Richard Swinburne. Oxford University Press, 1974.

TALLER Y LABORATORIO

THE "SALT-FOUNTAIN" AND THERMOHALINE CONVECTION. Melvin E. Stern en *Tellus: A Quarterly Journal of Geophysics,* vol. 12, n.º 2, págs. 172-175; mayo, 1960.

OBSERVATIONS OF THE CELL STRUCTURE OF SALT FINGERS. T. G. L. Shirtcliffe y J. S. Turner en *Journal of Fluid Mechanics,* vol, 41, parte 4, págs. 707-719; 15 de mayo de 1970.

Indice anual

Este índice contiene los nombres de los autores y artículos aparecidos en Investigación y Ciencia *a partir de enero de 1977. Se incorpora asimismo la relación alfabética de las secciones* Juegos matemáticos y Taller y laboratorio.

AUTORES

ARTICULOS

JUEGOS MATEMATICOS

TALLER Y LABORATORIO